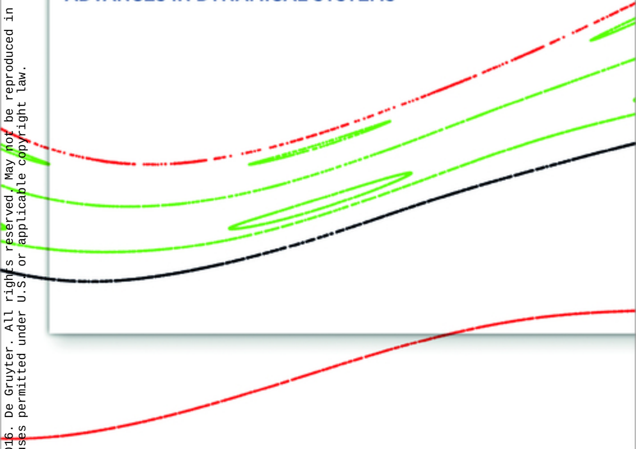
DE GRUYTER

Idris Assani (Ed.)

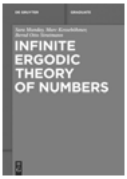
ERGODIC THEORY

ADVANCES IN DYNAMICAL SYSTEMS



Idris Assani (Ed.)
Ergodic Theory

Also of interest

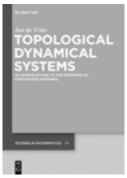


Infinite Ergodic Theory of Numbers
Sara Munday, Marc Kesseböhmer, Bernd Otto Stratmann, 2016
ISBN 978-3-11-043941-0, e-ISBN (PDF) 978-3-11-043942-7

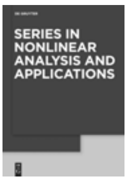
De Gruyter Studies in Mathematics
Carsten Carstensen et al. (Ed.)
ISSN 0179-0986

Volume 62
Positive Dynamical Systems in Discrete Time: Theory, Models, and Applications
Ulrich Krause, 2015
ISBN 978-3-11-036975-5, e-ISBN (PDF) 978-3-11-036569-6

Volume 61
Markov Operators, Positive Semigroups and Approximation Processes
Francesco Altomare, Mirella Cappelletti, Vita Leonessa, Ioan Rasa, 2014
ISBN 978-3-11-037274-8, e-ISBN (PDF) 978-3-11-036697-6



Volume 59
Topological Dynamical Systems: An Introduction to the Dynamics of Continuous Mappings
Jan Vries, 2014
ISBN 978-3-11-034073-0, e-ISBN (PDF) 978-3-11-034240-6



De Gruyter Series in Nonlinear Analysis and Applications
Jürgen Appell et al. (Ed.)
ISSN 0941-813X

Ergodic Theory

Advances in Dynamical Systems

Edited by
Idris Assani

DE GRUYTER

Mathematics subject classification 2010
54H20, 37A05, 60F05,37A25, 91A05, 60J20, 37J50, 70H99, 37C45, 54H20

Editor
Prof. Dr. Idris Assani
The University of North Carolina at Chapel Hill
Department of Mathematics
CB 3250 Phillips Hall
Chapel Hill NC 27599-3250
USA
assani@email.unc.edu

ISBN 978-3-11-046086-5
e-ISBN (PDF) 978-3-11-046151-0
e-ISBN (EPUB) 978-3-11-046091-9
Set-ISBN 978-3-11-046152-7

Library of Congress Cataloging-in-Publication Data
A CIP catalog record for this book has been applied for at the Library of Congress.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.dnb.de.

Cover image: Marie-Claude Arnaud
Typesetting: Lumina Datamatics
Printing and binding: CPI books GmbH, Leck
♾ Printed on acid-free paper
Printed in Germany

www.degruyter.com

Contents

Joseph Auslander and Gernot Greschonig*

Duality of the almost periodic and proximal relations

Abstract: We establish certain dualities between results on the almost periodic and the proximal relation of a compact flow. Moreover, we study the notions of distal and proximal equivalence of compact flows.

Keywords: Proximal, distal, minimal flow, almost periodic.

1 Introduction

A flow (X, T) is a left jointly continuous action of the topological group T on the compact Hausdorff space X. If (X, T) is a flow, a minimal set is a non-empty closed T invariant set M which is minimal with respect to these properties. By Zorn's Lemma, minimal sets always exist when X is compact. A non-empty subset M of X is minimal if and only if it is the orbit closure of each of its points – $\overline{Tx} = M$ for all $x \in M$. If $\overline{Tx}$ is minimal, x is said to be an almost periodic point. (X, T) is said to be a minimal flow if it is itself minimal so that $\overline{Tx} = X$ for all $x \in X$.

If (X, T) and (Y, T) are flows, a homomorphism from (X, T) to (Y, T) is a continuous onto map $\pi : X \to Y$ which intertwines the action of T, $\pi(tx) = t\pi(x)$ for $x \in X$ and $t \in T$. In this case we say that X is an extension of Y and Y is a factor of X. A homomorphism π determines a closed T invariant equivalence relation $R_\pi = \{(x, x')|\pi(x) = \pi(x')\}$, and every closed T invariant equivalence relation R on X determines a factor X/R. If M is a minimal set in X, then $\pi(M)$ is minimal in Y. In particular, a factor of a minimal flow is minimal.

The points x and x' are said to be proximal if there is a net $\{t_i\}$ in T such that $\lim t_i x = \lim t_i x'$, otherwise x and x' are said to be distal. We denote the proximal relation by P_X or P, so $(x, x') \in P_X$ if x and x' are proximal. The proximal relation is reflexive, symmetric and T invariant (if $(x, x') \in P_X$ and $t \in T$, then $(tx, tx') \in P_X$) but in general is not transitive or closed. A homomorphism $\pi : (X, T) \to (Y, T)$ of flows is proximal (respectively distal) if any two points $x, x' \in X$ with $\pi(x) = \pi(x')$ are proximal (respectively distal).

We say that $x \in X$ is a distal point if for all $x' \neq x$, x and x' are distal, and the flow (X, T) is distal of every point of X is distal, that is $P_X = \Delta_X$, the diagonal. A minimal

*Gernot Greschonig was supported by the 'Erwin-Schrödinger Fellowship' J3337 of the Austrian Science Fund FWF.

flow is said to be point distal if it contains a distal point. A result of Ellis [2] states that a minimal compact *metric* flow with a single distal point has a residual set of distal points.

If (X,T) is a flow, its enveloping semigroup $\mathcal{E}(X,T)$ (or $\mathcal{E}(X)$) is the closure of T in X^X with the product topology. It is easy to check that in fact $\mathcal{E}(X,T)$ is a semigroup under composition of maps for which multiplication is left continuous. If $\pi : X \to Y$ is a flow homomorphism, it induces a semigroup homomorphism $\theta : \mathcal{E}(X,T) \to \mathcal{E}(Y,T)$ such that $\pi(px) = \theta(p)\pi(x)$ for $x \in X$ and $p \in \mathcal{E}(X,T)$. We will suppress θ notationally and regard $\mathcal{E}(X,T)$ as acting on Y, writing $\pi(px) = p\pi(x)$.

Dynamical properties of the flow (X,T) are reflected in the algebraic properties of $\mathcal{E}(X,T)$. For example, x and x' are proximal if and only if $px = px'$ for all $p \in I$, a minimal left ideal in $\mathcal{E}(X,T)$ (i.e. a set with $\mathcal{E}(X,T) \cdot I \subset I$). A minimal left ideal contains idempotents (elements $u \in I$ with $uu = u$). We denote by $J(I)$ the set of idempotents in the minimal left ideal I. If u is an idempotent in $\mathcal{E}(X,T)$ and $x \in X$, then $(x,ux) \in P_X$. If u is a minimal idempotent (i.e. $u \in J(I)$ for some minimal left ideal I) and $x \in X$, then ux is an almost periodic point, and if x is an almost periodic point and I is a minimal left ideal in $\mathcal{E}(X,T)$, then $ux = x$ for some $u \in J(I)$. It follows that if x is a distal point, then $ux = x$ for all idempotents in $\mathcal{E}(X,T)$. In particular, $ux = x$ for all minimal idempotents, so a distal point is almost periodic.

2 Almost periodic and proximal relations of a flow

Definition 2.1. Let Ω_X be the almost periodic relation: $(x,x') \in \Omega_X$ if (x,x') is an almost periodic point in the product flow $(X \times X, T)$ with the diagonal action $t(x,x') = (tx,tx')$.

Clearly, $P_X \cap \Omega_X \subset \Delta_X$ with equality if (X,T) is minimal. By definition, $P_X = \Delta_X$ if and only if (X,T) is distal and $P_X = X \times X$ if and only if (X,T) is proximal. Dually, $\Omega_X = \Delta_X$ if and only if (X,T) is proximal (if x and x' are distal, then the orbit closure of (x,x') in $X \times X$ provides a minimal set disjoint from Δ_X), and $\Omega_X = X \times X$ if and only if (X,T) is distal. Moreover, if x is a distal point and x' is an almost periodic point, then $(x,x') \in \Omega_X$, since every minimal idempotent which fixes x' also fixes x.

Recall that if P_X is closed, it is an equivalence relation. It is not known whether this is the case for Ω_X. If the acting group T is *abelian* and the almost periodic points are dense in X, then Ω_X is dense in $X \times X$. Indeed, let $(x,x') \in X \times X$ and an open neighborhood U of (x,x') be arbitrary, and choose almost periodic points $y,y' \in X$ so that $(y,y') \in U$. If (z,z') is an almost periodic point in $\overline{T(y,y')}$, then the almost periodicity of y and y' implies that $y \in \overline{Tz}$ and $y' \in \overline{Tz'}$. Thus we can find $s,t \in T$ with $(sz,tz') \in U$, and since T is abelian, it follows from $(z,z') \in \Omega_X$ that $(sz,tz') \in \Omega_X$.

Therefore for minimal (X, T) with abelian T the relation Ω_X is closed if and only if it equals $X \times X$.

For a distal flow (X, T) with an arbitrary acting group T the diagonal action is distal on $X \times X$. Hence, every point in $X \times X$ is almost periodic and $\Omega_X = X \times X$ is an equivalence relation. In general, if the relation Ω_X is closed, then for every minimal idempotent $u \in \mathcal{E}(X, T)$ the set uX is closed. (Let $x' \in \overline{uX}$, and suppose $ux_n \to x'$. Then $(ux', ux_n) \in \Omega_X$, so $(ux', x') \in \Omega_X$. Since also $(ux', x') \in P_X$ we have $x' = ux' \in uX$.)

Definition 2.2. A finite subset $\{x_1, \dots, x_n\} \subset X$ is called a proximal set if there exists a net $\{t_i\} \subset T$ and a point $x \in X$ so that $t_i x_k \to x$ for $1 \leq k \leq n$. A subset B of X is called a proximal set if every finite subset of B is proximal.

A finite subset $\{x_1, \dots, x_n\}$ is called an almost periodic set if $(x_1, \dots, x_n)$ is an almost periodic point of the flow $\big(X^n, T\big)$, and $A \subset X$ is an almost periodic set if every finite subset of A is almost periodic. It follows that if $z \in X^A$ with range $z = A$, then z is an almost periodic point.

If B is a proximal set, any two points in B are proximal, but this says more. A corresponding statement holds for almost periodic sets.

Lemma 2.3. *Let B be a proximal set in X. Then there is a minimal left ideal I in $\mathcal{E}(X, T)$ such that if $r \in I$ the set rB is a singleton.*

Proof. If K is a finite subset of B, then there is a p_K in $\mathcal{E}(X, T)$ and $x^* \in X$ with $p_K K = \{x^*\}$. Regarding $\{p_K\}$ as a net, there is a subnet $\{p_{K_i}\}$ with $p_{K_i} \to p \in \mathcal{E}(X, T)$. Then if $x \in B$, $px = x^*$, and if $q \in \mathcal{E}(X, T)$, $qpx = qx^*$. It follows that there is a minimal left ideal with the desired properties. □

Lemma 2.4. *Let A be an almost periodic set in X. If I is a minimal left ideal in $\mathcal{E}(X, T)$, then there is a $u \in J(I)$ with $ux = x$ for all $x \in A$.*

Proof. Let $z \in X^A$ with range $z = A$. Since z is an almost periodic point in X^A, there is a $u \in J(I)$ with $uz = z$. Therefore $ux = x$ for all $x \in A$. □

A flow is distal if and only if every proximal set consists of a single point and is proximal if and only if every almost periodic set consists of a single point. A straightforward Zorn's Lemma argument shows that every proximal set is contained in a maximally proximal set, and similarly an almost periodic set is contained in a maximally almost periodic set.

Moreover, a singleton $\{x\}$ is a maximally proximal set if and only if x is a distal point. A singleton $\{x\}$ with x almost periodic is a maximally almost periodic set if and only if (X, T) is proximal, since if u is a minimal idempotent with $ux = x$ and $y \in X$ then $(x, uy) \in \Omega_X$ so $uy = x$.

Lemma 2.5. *If $\pi : (X, T) \to (Y, T)$ is a proximal extension, then every fibre $\pi^{-1}(y)$ is a proximal set. Similarly, if π is distal and $y \in Y$ is an almost periodic point, then the fibre $\pi^{-1}(y)$ is an almost periodic set.*

Proof. If π is proximal, $y \in Y$, and u is a minimal idempotent, then if $x, x' \in \pi^{-1}(y)$ we have $\pi(ux) = \pi(ux')$. Then $(ux, ux') \in P_X \cap \Omega_X$, so $ux = ux'$. Therefore $u\pi^{-1}(y)$ is a single point and $\pi^{-1}(y)$ is a proximal set.

Similarly, if π is distal and u is a minimal idempotent with $uy = y$, then ux and x are proximal and distal for all $x \in \pi^{-1}(y)$. Therefore $ux = x$ for all $x \in \pi^{-1}(y)$, so $\pi^{-1}(y)$ is an almost periodic set. □

Lemma 2.6. *Let (X, T) and (Y, T) be flows, and let $\pi : (X, T) \to (Y, T)$ be a homomorphism. Then $\pi(\Omega_X) = \Omega_Y$. If (Y, T) is pointwise almost periodic and $(y, y') \in P_Y$, then for every $x \in \pi^{-1}(y)$ there exists $x' \in \pi^{-1}(y')$ with $(x, x') \in P_X$, and hence $\pi(P_X) = P_Y$.*

Proof. We always have $\pi(\Omega_X) \subset \Omega_Y$ and $\pi(P_X) \subset P_Y$. Suppose that $(y, y') \in \Omega_Y$ and $\pi(x, x') = (y, y')$. Let u be a minimal idempotent in $\mathcal{E}(X, T)$ such that $(uy, uy') = (y, y')$. Then $\pi(ux, ux') = (uy, uy') = (y, y')$ and $(ux, ux') \in \Omega_X$.

If (Y, T) is pointwise almost periodic and $(y, y') \in P_Y$, then $y' = uy' = uy$ for some minimal idempotent u. Let $x \in X$ with $\pi(x) = y$. Then $\pi(ux) = uy = y'$ and $(x, ux) \in P_X$, since $ux = uux$. □

Example 2.7. The equality $\pi(P_X) = P_Y$ does not hold without almost periodicity of (Y, T). Let $X = [0, 1] \times \{0, 1\}$ be two copies of the unit interval, and let T be generated by the homeomorphism $(x, i) \mapsto (x^2, i)$ on X. Let Y be the quotient system obtained by identifying $(0, 0)$ with $(0, 1)$ and $(1, 0)$ with $(1, 1)$, and let $\pi : X \to Y$ be the homomorphism defined by this closed invariant equivalence relation. Then $(\pi(1/2, 0), \pi(1/2, 1))$ is in P_Y, but the unique inverse image $((1/2, 0), (1/2, 1))$ is not in P_X.

Corollary 2.8. *A factor of a distal flow is distal, and a factor of a proximal flow is proximal.*

Proof. Let $\pi : (X, T) \to (Y, T)$ be a homomorphism with (X, T) is distal. Then (X, T) is pointwise almost periodic, as is (Y, T), so $P_Y = \pi(P_X) = \pi(\Delta_X) = \Delta_Y$.

For a proximal flow, the proof is immediate. □

Lemma 2.9.

(i) *If $\pi : X \to Y$ is proximal, then $\pi^{-1}(P_Y) = P_X$.*
(ii) *If π is distal, then $\pi^{-1}(\Omega_Y) = \Omega_X$.*

Proof. (i): Always $\pi(P_X) \subset P_Y$, so $P_X \subset \pi^{-1}(P_Y)$. Let $(y, y') \in P_Y$ and $\pi(x, x') = (y, y')$, and let u be a minimal idempotent for which $uy = uy'$. Then $(ux, ux') \in \Omega_X$ and $\pi(ux) = \pi(ux')$. Since π is proximal, $(ux, ux') \in P_X$. But also $(ux, ux') \in \Omega_X$, so $ux = ux'$ and $(x, x') \in P_X$.

(ii): Let $(y, y') \in \Omega_Y$ and $\pi(x, x') = (y, y')$. Let u be a minimal idempotent such that $u(y, y') = (y, y')$. Then $(ux, x) \in P_X$ and $\pi(ux) = y = \pi(x)$. Since π is distal, $ux = x$. Similarly $ux' = x'$. Therefore $(x, x') = u(x, x') \in \Omega_X$. □

Using this lemma we can easily prove the following well-known result.

Theorem 2.10. *Let (X, T), (Y, T) and (Z, T) be compact flows, let $\alpha : (X, T) \to (Y, T)$ and $\beta : (Y, T) \to (Z, T)$ be homomorphisms and let $\pi = \beta \circ \alpha$.*
(i) *If α and β are distal, then π is distal. If (Z, T) is pointwise alomost periodic, the converse holds.*
(ii) *If α and β are proximal, then π is proximal.*

Proof. If α and β are distal, then clearly π is distal.

Suppose π is distal and (Z, T) is pointwise almost periodic. If $x, x' \in P_X$ with $\alpha(x) = \alpha(x')$, then $\pi(x) = \pi(x')$ and so $x = x'$. Now suppose $(y, y') \in P_Y$ with $\beta(y) = \beta(y')$. Let $x, x' \in X$ with $\alpha(x) = y$ and $\alpha(x') = y'$. Then $\pi(x) = \pi(x')$, and since (Z, T) is pointwise almost periodic $(\pi(x), \pi(x')) \in \Omega_Z$. From Lemma 2.9 (ii) follows $x, x' \in \Omega_X$, and then $(y, y') \in \Omega_Y \cap P_Y$ so $y = y'$.

Suppose π is proximal, and $\alpha(x) = \alpha(x')$. Then $\pi(x) = \pi(x')$, and $x, x' \in P_X$. If $\beta(y) = \beta(y')$, let $x, x' \in X$ with $\alpha(x) = y$ and $\alpha(x') = y'$. Then $\pi(x) = \pi(x')$ so $x, x' \in P_X$ and therefore $(y, y') \in P_Y$.

Now suppose α and β are proximal, and let $x, x' \in X$ with $\pi(x) = \pi(x')$. That is $\beta(\alpha(x)) = \beta(\alpha(x'))$, and $(\alpha(x), \alpha(x')) \in P_Y$. By Lemma 2.9, $x, x' \in P_X$. □

Definition 2.11. The proximal structure relation of a flow (X, T) is the smallest closed invariant equivalence relation on X so that the quotient flow is proximal. The distal structure relation of a flow (X, T) is the smallest closed invariant equivalence relation on X so that the quotient flow is distal.

Theorem 2.12. *The proximal structure relation of a flow (X, T) is the closed invariant equivalence relation generated by Ω_X. For a pointwise almost periodic flow (X, T) the relation Ω_X is a closed invariant equivalence relation if and only if X is a distal extension of a proximal flow.*

The distal structure relation of (X, T) is the closed invariant equivalence relation generated by P_X. The relation P_X is closed invariant if and only if X is a proximal extension of a distal minimal flow.

Proof. If R is a closed invariant equivalence relation containing Ω_X and $Y = X/R$ with $\pi : (X, T) \to (Y, T)$, then (since $\pi(\Omega_X) = \Omega_Y$ by Lemma 2.6), $\Omega_Y = \Delta_Y$ so (Y, T) is proximal.

Now suppose that (X, T) is pointwise almost periodic, $\pi : (X, T) \to (Y, T)$ is distal and (Y, T) is proximal. Let (x_1, x_2) and $(x_2, x_3) \in \Omega_X$. Then $(\pi(x_1), \pi(x_2)) \in \Omega_Y$ and since (Y, T) is proximal $\pi(x_1) = \pi(x_2)$. Similarly $\pi(x_2) = \pi(x_3)$, so $\pi(x_1) = \pi(x_3)$. Since π is distal and (Y, T) is pointwise almost periodic, by Lemma 2.5 $(x_1, x_3) \in \Omega_X$ so that Ω_X is an equivalence relation. If $(x, x') \in \overline{\Omega_X}$, then $\pi(x) = \pi(x')$ and again since π is distal $x, x' \in \Omega_X$. Hence Ω_X is closed, and the T-invariance is immediate.

Suppose $\pi : (X, T) \to (Y, T)$, with $\pi(P_X) \subset \Delta_Y$. We want to show that Y is distal. Let $y \in Y$ and let u be a minimal idempotent in $\mathcal{E}(X, T)$. Then $(x, ux) \in P_X$ so $y = \pi(x) = \pi(ux) = uy$. This shows that Y is pointwise almost periodic and every point in Y is a distal point. The (well-known) assertions for P_X and the distal structure relation can be verified in duality to the arguments above. □

This corollary can be easily relativised.

Theorem 2.13. *Let $\pi : (X, T) \to (Y, T)$ be a homomorphism of minimal flows. Let $R_\pi = \{(x, x') : \pi(x) = \pi(x')\}$ be the relation of π, and let $P_\pi = P_X \cap R_\pi$ and $\Omega_\pi = \Omega_X \cap R_\pi$. Let R_P and R_Ω denote the closed invariant equivalence relations generated by P_π and Ω_π respectively. Then the induced homomorphisms from X/R_P to Y and X/R_Ω to Y are distal and proximal respectively.*

3 Distal and proximal equivalence of minimal flows

Throughout this section we will restrict our considerations to compact minimal flows.

Definition 3.1. The minimal flows (X, T) and (Y, T) are said to be distally equivalent if they have a common distal extension. That is, there is a minimal flow (Z, T) and distal homomorphisms $\alpha : (Z, T) \to (X, T)$ and $\beta : (Z, T) \to (Y, T)$. Similarly, (X, T) and (Y, T) are said to be proximally equivalent if they have a common proximal extension.

Proposition 3.2. *Distal and proximal equivalence are equivalence relations. The distal equivalence class of a distal minimal flow consists of the family of distal minimal flows, and the corresponding statement holds for the proximal equivalence class of a proximal minimal flow.*

Proof. Both relations are obviously reflexive and symmetric. The proof of transitivity uses the joining of two flows. Let (X, T), (Y, T) and (Z, T) be flows. Let (Q, T) be a common distal extension of $(X, T) = \alpha(Q, T)$ and $(Y, T) = \beta(Q, T)$, and let (R, T) be a common distal extension of $(Y, T) = \gamma(R, T)$ and $(Z, T) = \delta(R, T)$. We choose a point $(q, r) \in Q \times R$ with $\beta(q) = \gamma(r)$, and let (S, T) be a minimal set in the T-orbit closure of

(q, r). Then (S, T) is a common extension of (X, T) and (Z, T) with the homomorphisms $\alpha \circ \pi_1$ and $\delta \circ \pi_2$, respectively. If $\alpha \circ \pi_1(q', r') = \alpha \circ \pi_1(q'', r'')$ for $(q', r'), (q'', r'') \in S$, then either $q' \neq q''$ or $\beta(q') = \beta(q'') = \gamma(r') = \gamma(r'')$. If $q' \neq q''$ with $\alpha(q') = \alpha(q'')$, then (q', r') and (q'', r'') are distal in (S, T) by the distality of α. Otherwise (q', r') and (q'', r'') with $\gamma(r') = \gamma(r'')$ are distal in (S, T) by the distality of γ. Hence $\alpha \circ \pi_1$ is a distal homomorphism, as well as $\delta \circ \pi_2$.

The proof of the transitivity of proximal equivalence follows in duality. If $\alpha \circ \pi_1(q', r') = \alpha \circ \pi_1(q'', r'')$ for $(q', r'), (q'', r'') \in S$, then by the proximality of α there exists a point $((q, r), (q, \bar{r}))$ in the T-orbit closure of $((q', r'), (q'', r''))$ in $S \times S$. Since $\gamma(r) = \gamma(\bar{r})$, the proximality of γ implies that (q', r') and (q'', r'') are proximal.

Given two arbitrary distal flows (Q, T) and (R, T), the joining (S, T) is a common distal extension, dually for two proximal flows. The statement on the distal (proximal) equivalence class of a distal (proximal) flow follows. □

Lemma 3.3. *Let (X, T) and (Y, T) be minimal flows, and let $\pi : (X, T) \to (Y, T)$ be a homomorphism. Let $(y, y') \in P_Y$ and suppose $x \in X$ with $\pi(x) = y$. Then there is an $x' \in \pi^{-1}(y')$ with $(x, x') \in P_X$.*

Proof. Let I be a minimal left ideal in $\mathcal{E}(X, T)$ such that $py = py'$ for $p \in I$. Let u be an idempotent in I for which $y' = uy$. Then $(x, ux) \in P_X$ and $\pi(x, ux) = (y, uy) = (y, y')$. □

Theorem 3.4. *For minimal flows we have the following assertions:*

(i) *'P is an equivalence relation', 'P is closed' and 'P is dense' are preserved under factors and proximal equivalence.*

(ii) *'Ω is an equivalence relation' is preserved under distal equivalence.*

(iii) *'Ω is closed' and 'Ω is dense' are preserved under factors and distal equivalence.*

Proof. Let $\pi : (X, T) \to (Y, T)$. Since $\pi(P_X) = P_Y$ and $\pi(\Omega_X) = \Omega_Y$ by Lemma 2.6, the property of P being closed or dense is preserved under factors, and similarly for Ω.

The preservation of the property of P and Ω being an equivalence relation, under respectively factors and distal homomorphisms, follows from the preceding lemma and Lemma 2.9.

Suppose π is distal, and Ω_Y is dense in $Y \times Y$. Let $(x, x') \in X \times X$ and $(y, y') = \pi(x, x')$. Let $(y_n, y'_n) \to (y, y')$ with $(y_n, y'_n) \in \Omega_Y$. Since a distal homomorphism is open, there are $(x_n, x'_n) \to (x, x')$ with $\pi(x_n, x'_n) = (y_n, y'_n)$, and since π is distal $(x_n, x'_n) \in \Omega_X$ by Lemma 2.9.

The lifting of the density of P under proximal extensions follows from the 'semi-openness' of extensions of minimal flows. If $\pi : (X, T) \to (Y, T)$ is a proximal homomorphism of minimal flows with P_Y dense in Y, let U_1 and U_2 be non-empty open sets in X. Let $V_i = \operatorname{int} \pi(U_i)$ $(i = 1, 2)$. Let $(y_1, y_2) \in V_1 \times V_2 \cap P_Y$, and let $(x_1, x_2) \in U_1 \times U_2$

with $\pi(x_1, x_2) = (y_1, y_2)$. Since π is proximal, $(x_1, x_2) \in P_X$ by Lemma 2.9. (This proof is from [3], where in fact a more general result is obtained.) $\square$

In general, 'Ω an equivalence relation' is not preserved under factors. In fact, if (M, T) is the universal minimal flow, Ω_M is an equivalence relation, and every minimal flow (X, T) is a factor of (M, T).

'Point distal' is preserved under distal equivalence. For, if $\pi : (X, T) \to (Y, T)$ is a homomorphism and $x \in X$ is a distal point, then $y = \pi(x)$ is a distal point. If π is distal and $y \in Y$ is a distal point, then all $x \in \pi^{-1}(y)$ are distal points.

Theorem 3.5. *Let A be an almost periodic set in X, let $z \in X^A$ with range $z = A$, and let $w \in \overline{Tz}$. Then $B =$ range w is an almost periodic set. If A is maximal, so is B.*

Proof. Since $\overline{Tz}$ is a minimal set, all of its points are almost periodic, so B is an almost periodic set. Suppose A is maximal, and suppose $\{x\} \cup B$ is almost periodic. Let x' such that $(z, x') \in \overline{T(w, x)}$. Then $A \cup \{x'\}$ is an almost periodic set, so by maximality of A, $x' \in A$. It follows that $x \in B$, so B is maximal. $\square$

If (X, T) is a compact flow and $x \in X$, then the capturing set is defined by

$$C(x) = \{x' \in X : x \in \overline{Tx'}\}.$$

We have in duality to the above theorem the following result.

Theorem 3.6. *If B is a proximal set, $z \in X^B$ with range $z = B$ and $w \in C(z)$, then range w is a proximal set.*

Proof. If $\{x_1, \ldots, x_n\}$ be a finite subset of range w and t_k is a net in T with $t_k w \to z$, then the set $\{y_1, \ldots, y_n\} \subset B$ with $y_i = \lim t_k x_i$ is a finite proximal set. It follows immediately that $\{x_1, \ldots, x_n\}$ is proximal, as well as range w. $\square$

Invariants for distal equivalence are obtained by minimal idempotents. Let I be a minimal left ideal in $\mathcal{E}(X, T)$, and let $J(I)$ denote the idempotents in I. If $x \in X$, let $J(x) = \{u \in J(I) : ux = x\}$.

Lemma 3.7. *Let (X, T) and (Y, T) be minimal flows, and let $x \in X$, $y \in Y$. Then the following are equivalent:*

(i) *There is a minimal flow (X', T), an $x' \in X'$, homomorphisms $\pi : X' \to X$ and $\sigma : X' \to Y$, with π distal, $\pi(x') = x$ and $\sigma(x') = y$.*
(ii) *$J(px) \subset J(py)$ for all $p \in I$.*

Proof. (i) $\implies$ (ii): If $u \in J(px)$, then $\pi(upx') = \pi(px')$, so that by the distality of π and Lemma 2.5 $(upx', px') \in \Omega_{X'}$. Since also $(upx', px') \in P_{X'}$, it follows that $upx' = px'$. Hence $u \in J(px')$ and therefore $u \in J(py)$.

(ii) $\implies$ (i): Let $x' = (x, y)$ and let $X' = \overline{Tx'}$. Since $J(x) \subset J(y)$, x' is an almost periodic point, so (X', T) is a minimal flow. Let π and σ be (the restrictions of) the projections from X' to X and Y respectively. Suppose $\pi(px, py) = \pi(qx, qy)$ with $p, q \in \mathcal{E}(X', T)$ so that $px = qx$, and let $u \in \mathcal{E}(X', T)$ be an idempotent so that $u(px, py) = (px, py)$ and $u(qx, qy) = (px, py)$. Then $uqx = upx = px = qx$ and $upy = py$ and $uqy = qy = py$. Thus $(px, py) = (qx, qy)$ and so π is distal. □

Theorem 3.8. *The minimal flows (X, T) and (Y, T) are distally equivalent if and only if there are $x \in X$ and $y \in Y$ such that $J(px) = J(py)$ for all $p \in I$.*

Proof. If $J(px) = J(py)$ for all $p \in I$, then the projections π and σ from X' to X and Y, respectively, as in the lemma are both distal, so the flows are distally equivalent.

If (X, T) and (Y, T) are distally equivalent, then there is a minimal flow (Z, T) and distal homomorphisms $\pi : Z \to x$ and $\psi : Z \to Y$. Let $z \in Z$ and $x = \pi(z)$, $y = \psi(z)$. If $p \in I$ and $u \in J(px)$, then $\pi(upz) = upx = px = \pi(pz)$. Since π is distal, $upz = pz$, and so $upy = py$, and $u \in J(py)$. That is, $J(px) \subset J(py)$ and by symmetry $J(py) \subset J(px)$. □

Theorem 3.9. *Let (X, T) be a minimal flow.*

(i) *Let I be a fixed minimal left ideal in $\mathcal{E}(X, T)$. Then if $u \in J(I)$, uX is a maximal almost periodic set and all maximal almost periodic sets are of this form.*

(ii) *Let x be a point in X and I be a minimal ideal in $\mathcal{E}(X, T)$. Then the set $B_I(x) = \{x' \in X : px' = px$ for every $p \in I\}$ is a maximal proximal set and all maximal proximal sets are of this form. For every $u \in J(I)$ the set $B_u(x) = \{x' \in X : ux' = ux\}$ coincides with $B_I(x)$.*

Proof. (i): Clearly, uX is an almost periodic set. If $\{x'\} \cup uX$ is an almost periodic set, then by Lemma 2.4 there exists a $v \in J(I)$ with $vx' = x'$ and $vux' = ux'$. From $(x', ux') \in P_X \cap \Omega_X$, it follows that $x' = vx' = ux'$. Therefore uX is maximally almost periodic.

Now suppose A is maximally almost periodic. Let I be a minimal left ideal in $\mathcal{E}(X, T)$, and by Lemma 2.4 let $u \in J(I)$ such that $ux = x$ for all $x \in A$. We first note that $uX \subset A$. For if $x' \in uX$, then $A \cup x'$ is an almost periodic set, so by maximality of A, $x' \in A$. Now suppose $y \in A$. Then $uy \in uX \subset A$, so we have $(y, uy) \in \Omega_X \cap P_X$, and therefore $y = uy \subset uX$. That is, $A = uX$.

(ii): Clearly, $B_I(x)$ is a proximal set. By Lemma 2.3 every proximal set B in X is subset of $B_I(x)$ for some minimal ideal I in $\mathcal{E}(X, T)$, and by the minimality of the ideal I the set $B_I(x)$ is a maximal proximal set. Since the set $B_u(x)$ is a proximal set and two minimal ideals are either disjoint or coincide, the sets $B_I(x)$ and $B_u(x)$ coincide. □

Note that if A is an almost periodic set and B is a proximal set, then $|A \cap B| \leq 1$. Now let A be a maximal almost periodic set and B be a maximal proximal set. We show that $|A \cap B| = 1$. Let I be a minimal left ideal and $x \in X$ with $B = B_I(x)$, and let $u \in J(I)$ with $A = uX$. Then $ux \in A$ and $ux \in B_u(x) = B_I(x)$, so $A \cap B = \{ux\}$.

If I is a minimal ideal in $\mathcal{E}(X,T)$ and $u \in J(I)$, then uI is a group with identity u (cf. [1]). Moreover, the sets uI for $u \in J(I)$ define a partition of I into subgroups which are all isomorphic. This group $\mathcal{G}(X,T)$ is called the Ellis group of (X,T), and for $x \in X$ with $ux = x$ we define a subgroup $\mathcal{G}(X,x) = \{p \in uI : px = x\}$. From $Ix = X$ and Theorem 3.9, it follows that $\mathcal{G}(X,T)x$ is a maximal almost periodic set.

Theorem 3.10. *Let (X,T) and (Y,T) be minimal flows with $x \in X$, $y \in Y$. Then there is a homomorphism $\sigma : (X,T) \to (Y,T)$ with $\sigma(x) = y$ if and only if $\mathcal{G}(X,x) \subset \mathcal{G}(Y,y)$ and $J(px) \subset J(py)$ for all $p \in I$. σ is an isomorphism if and only if $\mathcal{G}(X,x) = \mathcal{G}(Y,y)$ and $J(px) = J(py)$ for all $p \in I$.*

Proof. In the notation of Lemma 3.7, since $\mathcal{G}(X,T)$ and $\mathcal{G}(Y,T)$ are homomorphic images of $\mathcal{G}(X',T)$ and $\mathcal{G}(X,x) \subset \mathcal{G}(Y,y)$, it follows that $\mathcal{G}(X',x') = \mathcal{G}(X,x)$. Since a maximal almost periodic set in X' is mapped one to one onto a maximal almost periodic set in X, the distal homomorphism π is an isomorphism. The reverse implication is obvious. □

Corollary 3.11. *Let the minimal flows (X,T) and (Y,T) be distally and proximally equivalent. Then (X,T) and (Y,T) are isomorphic.*

Proof. For distally equivalent flows, it follows from Theorem 3.8 that there are $x \in X$ and $y \in Y$ with $J(px) = J(py)$ for every $p \in I$. Moreover, for a proximal extension $\tau : (Z,T) \to (X,T)$ with $\tau(z) = x$, it follows that $\mathcal{G}(X,x) = \mathcal{G}(Z,z)$. Indeed, the orbits under $\mathcal{G}(X,T)$ and $\mathcal{G}(Z,T)$ are maximal almost periodic sets in X and Z respectively, and every fibre of τ can intersect a maximal almost periodic set in Z in at most one point. □

In the following results we will use the so-called *circle operation* of the enveloping semigroup on the set of subsets of X. Let $p \in \mathcal{E}(X,T)$, let A be a subset of X and let t_i be a net in T with $t_i \to p$. Then $p \circ A$ consists of all points $x \in X$ so that there is a net x_i in X with $t_i x_i \to x$. It follows easily for $p, q \in \mathcal{E}(X,T)$ and $A \subset X$ that $p \circ (q \circ A) = (pq) \circ A$.

Theorem 3.12. *If A is a maximal almost periodic set, then the orbit closure of $u \circ A$ in 2^X is a proximal minimal flow.*

Proof. Recall that $A = uX$ for some minimal idempotent u. It is sufficient to show that if $p \in I$, $up \circ uX = u \circ uX$.

Note: If $q \in I$ and $u \in J(I)$, then $uquX = uX$. Since uI is a group with identity u, there exists $p \in I$ with $uquup = uqup = u$. Then $uX = uqupX \subset uquX$, while the inclusion $uquX \subset uX$ is trivial. Hence $u \circ uX = u \circ upuX \subset u \circ up \circ uX = uup \circ uX = up \circ uX$ and also $up \circ uX = uup \circ uX \subset u \circ up \circ uX \subset u \circ uX$ (since $p \circ uX \subset X$). □

Theorem 3.13. *Let (M, T) be the universal minimal flow for T.*

The following are equivalent:

(i) *Ω_M is dense in $M \times M$.*

(ii) *There are no non-trivial minimal proximal flows (X, T).*

(iii) *If A is a maximal almost periodic set in M, then $u \circ A = M$.*

Proof. (i) $\Longrightarrow$ (ii): If Ω_M is dense in $M \times M$, then Ω_X is dense in $X \times X$, so (X, T) is not proximal. Therefore (i) $\Longrightarrow$ (ii).

(ii) $\Longrightarrow$ (iii): If $u \circ A$ is a proper subset of M, then by Theorem 3.12 its orbit closure is a non-trivial proximal flow.

(iii) $\Longrightarrow$ (i): Suppose $A = uX$, and let $m \in A$. Let $\{t_j\}$ be a net in T with $t_j \to u$. Let $m' \in M$ and let $m_j \in A$ such that $t_j m_j \to m'$. Now $(t_j m_j, t_j m) \in \Omega_M$ and it follows that $(m', m) \in \overline{\Omega_M}$. Therefore $M \times \{m\} \in \overline{\Omega_M}$, so $M \times M \subset \overline{\Omega_M}$. □

Theorem 3.14.

(i) *Let A be a maximal almost periodic set, and let $x \in X$. Then there is an $x' \in A$ such that $(x, x') \in P_X$.*

(ii) *Let B be a maximal proximal set, and let $x \in X$. Then there is an $x' \in B$ such that $(x, x') \in \Omega$.*

Proof. (i): Let $z \in X^A$ with range $z = A$. Let $u \in J(I)$ with $uz = z$ and let $x' = ux$. Then $A \cup \{x'\}$ is an almost periodic set. By maximality of A, $x' \in A$ and $(x, x') \in P_X$.

(ii): There is a minimal left ideal I, $u \in J(I)$ with $ux = x$, and $y \in X$ such that $B = B_I(y) = B_u(y)$. Then, if $x' = uy$, $(x, x') \in \Omega$ and $x' \in B_u(y)$. □

4 Proximal and almost periodic cells of a flow

The proximal cell of $x \in X$ is defined by $P_X(x) = \{x' \in X : (x, x') \in P_X\}$. Note that unless P_X is an equivalence relation, it is in general not the case that if x' and x'' are in $P_X(x)$ then $(x', x'') \in P_X$. It is well known that P_X is an equivalence relation if and only if there exists a unique minimal left ideal in $\mathcal{E}(X, T)$.

In duality the almost periodic cell is defined by $\Omega_X(x) = \{x' \in X : (x, x') \in \Omega_X\}$. Dually, Ω_X is an equivalence relation if and only if given a minimal left ideal I in $\mathcal{E}(X, T)$ and $x \in X$ and there is a unique idempotent (if any) in I with $ux = x$. Suppose $ux = vx$ with $u \neq v$. Then there is a $y \in X$ with $uy \neq vy$ so that $(x, uy) \in \Omega_X$, $(x, vy) \in \Omega_X$ and $(uy, vy) \notin \Omega_X$. Conversely, if the 'unique idempotent' condition holds, $(x,y),(y,z) \in \Omega_X$, $u(x,y) = (x,y)$ and $v(y,z) = (y,z)$, then $uy = vy = y$ so $u = v$ and $u(x,z) = (x,z)$.

Theorem 4.1. *The following are equivalent:*

(i) *P_X is an equivalence relation.*
(ii) *If B is a proximal set, $z \in X^B$ with range $z = B$ and $w \in \overline{Tz}$, then range w is a proximal set.*
(iii) *Every proximal cell is a proximal set.*
(iv) *Every proximal cell is a maximal proximal set.*

Proof. Suppose P_X is an equivalence relation, and B, z and w are as in (ii). We first show that the only minimal set in $\overline{Tz}$ is the set of elements in X^B with a trivial (singleton) range. Let N be such a minimal set. Then there is a minimal left ideal I in $\mathcal{E}(X,T)$ such that $N = \{pz : p \in I\}$. Let $x, x' \in B$ and let $u \in J(I)$. Then $(x,ux) \in P_X$ and $(x',ux') \in P_X$. Since x and x' are in the proximal set B, also $(x,x') \in P_X$. Since P_X is an equivalence relation, $(ux,ux') \in P_X$. Since also $(ux,ux') \in \Omega_X$, $ux' = ux$. That is, range $uz = \{ux\}$, so every element of N has a trivial range.

Now let K be a minimal set in $\overline{Tw}$. Since $w \in \overline{Tz}$ we have $K \subset \overline{Tz}$, so K consists of elements of X^B with a trivial range. Therefore range w is a proximal set.

(ii) $\Longrightarrow$ (iii): If u is a minimal idempotent and $(x,x') \in P_X$, then $(ux,ux') \in P_X$. But also $(ux,ux') \in \Omega_X$, so $ux = ux'$. That is, if $x' \in P_X(x)$, then $ux' = ux$ so that $P_X(x)$ is a proximal set.

(iii) $\Longrightarrow$ (i): Let $(x,x') \in P_X$ and $(x',x'') \in P_X$. Then x and x'' are in $P_X(x')$ which is a proximal set, so $(x,x'') \in P_X$. Therefore P_X is an equivalence relation.

The equivalence of (iii) and (iv) is obvious. □

Theorem 4.2. *Let (X,T) be a minimal flow. Then the following are equivalent:*

(i) *Ω_X is an equivalence relation.*
(ii) *If I is a minimal left ideal in $\mathcal{E}(X,T)$ and u and v are idempotents in I with $ux = vx$ for some $x \in X$, then $uy = vy$ for all $y \in X$.*
(iii) *Every almost periodic cell is an almost periodic set.*
(iv) *Every almost periodic cell is a maximal almost periodic set.*

Proof. (i) $\Longrightarrow$ (ii): Suppose Ω_X is an equivalence relation and $ux = vx$. Let $y \in X$. Now $(uy,ux) \in \Omega_X$ and $(ux,vy) = (vx,vy) \in \Omega_X$. Hence $(uy,vy) \in \Omega_X$. But also $(uy,vy) \in P_X$, so $uy = vy$.

(ii) $\Longrightarrow$ (iii): Let $x \in X$ and let $u \in J(I)$ with $ux = x$. If $y \in \Omega_X(x)$, then there exist a $v \in J(I)$ so that $v(x,y) = (x,y)$. Since $ux = vx$, it follows by the assumption $uy = vy$, so that $uy = y$. Therefore $\Omega_X(x)$ is an almost periodic set.

(iii) $\Longrightarrow$ (i): Suppose $(x,x') \in \Omega_X$ and $(x',x'') \in \Omega_X$. Then x and x'' are in $\Omega_X(x')$ which is an almost periodic set, so $(x,x'') \in \Omega_X$. Therefore Ω_X is an equivalence relation.

The equivalence of (iii) and (iv) is obvious. □

References

[1] R. Ellis, A semigroup associated with a transformation group, *Trans. Amer. Math. Soc.* **94** (1960), 272–281.

[2] R. Ellis, The Veech structure theorem, *Trans. Amer. Math. Soc.* **186** (1973), 203–218.

[3] E. Glasner, Topological weak mixing and quasi-Bohr systems, *Israel J. Math.* **148** (2005), 277–304.

Jean-Pierre Conze and Stéphane Le Borgne

Limit directions of a vector cocycle, remarks and examples

Abstract: We study the set $\mathcal{D}(\Phi)$ of limit directions of a vector cocycle (Φ_n) over a dynamical system, i.e., the set of limit values of $\Phi_n(x)/\|\Phi_n(x)\|$ along subsequences such that $\|\Phi_n(x)\|$ tends to ∞. This notion is natural in geometrical models of dynamical systems where the phase space is fibred over a basis with fibers isomorphic to $\mathbb{R}^d$, like systems associated to the billiard in the plane with periodic obstacles. It has a meaning for transient or recurrent cocycles.

Our aim is to present some results in a general context as well as for specific models for which the set of limit directions can be described. In particular we study the related question of sojourn in cones of the cocycle when the invariance principle is satisfied.

Keywords: dynamical system, vector cocycle, essential values, limit directions, invariance principle.

Introduction

Let (X, μ, T) be an ergodic dynamical system and Φ be a measurable function on X with values in $\mathbb{R}^d$. The ergodic sums $\Phi_n(x) := \sum_{k=0}^{n-1} \Phi(T^k x), n \geq 1$, define a vector process. When Φ is integrable and not centered, this process tends a.s. to ∞ in the direction of the mean $\int \Phi \, d\mu$. A general question, when Φ is centered or for a measurable non integrable Φ, is to find in which directions at infinity the ergodic sums are going. The set of these directions is a kind of boundary for the cocycle (Φ_n), i.e., for the process of ergodic sums.

This leads to the notion of limit directions and to the cohomologically invariant notion of essential limit directions. The limit directions of a vector cocycle (Φ_n) over a dynamical system can be defined as the limit values of $\Phi_n(x)/\|\Phi_n(x)\|$ along subsequences such that $\|\Phi_n(x)\|$ tends to ∞.

The notion of limit directions is natural in geometrical models of dynamical systems where the phase space is fibred over a basis with fibers isomorphic to $\mathbb{R}^d$, like the dynamical systems associated to the billiard in the plane with periodic obstacles. It has a meaning for recurrent cocycles as well as for transient cocycles.

Our aim is to present some results in a general context (Section 2) and for specific models where the set of limit directions can be made explicit. In Subsection 2.5 1-dimensional cocycles are considered and some classical results are recalled or slightly extended.

In Section 3.1, we apply properties like the CLT for subsequences or the invariance principle to study essential limit directions and the behavior of the process induced by the cocycle on the sphere. For $d \geq 2$, when Φ satisfies a Central Limit Theorem, one can think that the limit behavior of the sums is analogous to that of a Brownian motion, in particular in terms of visit of cones. In the last subsection 3.2 this is shown to be the case, at least if Φ satisfies Donsker's invariance principle.

1 Preliminaries

Let $(X, \mathcal{B}, \mu, T)$ be a dynamical system, where $(X, \mathcal{B})$ is a standard Borel space, T an invertible measurable map $T : X \to X$ and μ a probability measure which is T-invariant. Let Φ be a measurable function on X with values in $G = \mathbb{R}^d$. The process $(\Phi \circ T^n)_{n \geq 1}$ is stationary. Recall that, under general assumptions, every stationary process (X_n) with values in $\mathbb{R}^d$ can be represented in such a way for some dynamical system and some measurable Φ.

Part of the results below are valid when μ is only supposed to be a σ-finite T-quasi-invariant measure such that T is *conservative* for μ, i.e., for every measurable B in X, for μ-a.e. $x \in B$, there is $n(x) > 0$ such that $T^{n(x)}x \in B$. Nevertheless for the sake of simplicity, excepted in Subsection 2.5, we will restrict the presentation to the framework of a probability invariant measure μ.

Excepted in Section 2.5, the system (X, μ, T) is supposed to be ergodic. All equalities are understood to hold μ-a.e. All sets that we consider are measurable and (unless the contrary is explicitly stated) with positive measure.

To Φ is associated a *cocycle* $(\Phi_n)_{n \in \mathbb{N}}$ defined by $\Phi_0(x) = 0$,

$$\Phi_n(x) = \Phi(x) + \ldots + \Phi\left(T^{n-1}x\right), \text{ for } n \geq 1, \text{ and } \Phi_n = -\Phi_{-n} \circ T^n, \text{ for } n < 0,$$

and a map T_Φ (called *skew product*) acting on $X \times \mathbb{R}^d$ by

$$T_\Phi : (x, y) \to \left(Tx, y + \Phi(x)\right). \tag{1.1}$$

The cocycle relation $\Phi_{n+p}(x) = \Phi_n(x) + \Phi_p(T^n x), \forall n, p \in \mathbb{Z}$ is satisfied. The cocycle gives the position in the fiber after n iterations of T_Φ:

$$T_\Phi^n(x, y) = \left(T^n x,\ y + \Phi_n(x)\right).$$

The cocycle (Φ_n) can be viewed as a *"stationary" walk* in $\mathbb{R}^d$ "driven" by the dynamical system (X, μ, T). It is also the sequence of ergodic sums of Φ for the action of T. We will use as well the notation (Φ, T).

The Lebesgue measure on $\mathbb{R}^d$ is denoted by $m(dy)$ or simply dy. The map T_Φ leaves invariant the measure $\mu \times m$ denoted by λ.

Recall that a cocycle (Φ_n) over (X, μ, T) is *transient* if $\lim_n \|\Phi_n(x)\| = +\infty$, for a.e. $x \in X$. It is recurrent if $\liminf_n \|\Phi_n(x)\| < \infty$, for a.e. $x \in X$. It is well known that, when T is ergodic, a cocycle is either transient or recurrent (see the comment below).

Recurrence of the cocycle is equivalent to conservativity of the map T_Φ for the measure λ. When (Φ_n) is recurrent, then $(\Phi_n(x))$ returns for a.e. x infinitely often in any neighborhood of the origin. In dimension 1, if Φ is integrable and (X, μ, T) is ergodic, then $(\Phi_n)_{n\in\mathbb{Z}}$ is recurrent if and only if $\mu(\Phi) = 0$. In higher dimension, recurrence requires stronger assumptions.

Induced map

Let us recall some definitions and notations about induced maps.

Let B be a measurable set of positive μ-measure. On B equipped with the measure $\mu_B = \mu(B)^{-1}\mu_{|B}$, the induced transformation is $T_B(x) = T^{R(x)}(x)$, where $R(x)$ is the return time $R(x) = R_B(x) := \inf\{j \geq 1 : T^j x \in B\}$. The return time is well defined for a.e. $x \in B$ by conservativity of systems with finite invariant measure. We induce[1] Φ on B by putting

$$\Phi^B(x) := \Phi_{R(x)}(x) = \sum_{j=0}^{R(x)-1} \Phi\left(T^j x\right). \tag{1.2}$$

The "induced" cocycle is $\Phi_n^B(x) := \Phi^B(x) + \Phi^B(T_B x) \cdots + \Phi^B(T_B^{n-1}x)$, for $n \geq 1$.

If Φ is recurrent, then each induced cocycle (Φ_n^B) is recurrent. Indeed $(T_B)_{\Phi_B}$ is the induced map on $B \times G$ of T_Φ which is conservative.

If T is ergodic, then (B, μ_B, T_B) is ergodic. The converse is true when $X = \bigcup_n T^n B$.

When the map T is ergodic, the above formulas can be extended to X by setting, for every measurable set B of positive measure, for a.e. $x \in X$:

$$R_B(x) = \inf\left\{j \geq 1 : T^j x \in B\right\}, \quad \Phi^B(x) = \sum_{j=0}^{R_B(x)-1} \Phi\left(T^j x\right). \tag{1.3}$$

Recall that two $\mathbb{R}^d$-valued cocycles (Φ^1, T) and (Φ^2, T) over the dynamical system (X, μ, T) are *μ-cohomologous* with *transfer function* Ψ, if there is a measurable map $\Psi : X \to \mathbb{R}^d$ such that

$$\Phi^1(x) = \Phi^2(x) + \Psi(Tx) - \Psi(x), \text{ a.e.} \tag{1.4}$$

Φ is a *μ-coboundary*, if it is cohomologous to 0.

We choose a norm $\|.\|$ on $\mathbb{R}^d$. We will use the inequality

$$\left|\left\|\Phi_{n+1}(x)\right\| - \left\|\Phi_n(Tx)\right\|\right| \leq \left\|\Phi(x)\right\|. \tag{1.5}$$

1 In short the function Φ itself will also be called "cocycle" and Φ^B "induced cocycle" on B.

2 Limit directions of a vector cocycle, general properties

“0 -1” properties for a cocycle

Let (Φ_n) be a cocycle over an ergodic dynamical system (X, μ, T). Some of its limit properties are related to the ergodicity of the skew product T_Φ. For example, equirepartition properties (comparison of the number of visits to sets of finite measure) are given by the ratio ergodic theorem when the skew product T_Φ is ergodic.

There are also limit properties which do not a priori require ergodicity of the skew product, but appear as “0 -1” properties, in the sense that either they are satisfied by a.e. x, or are not satisfied by a.e. x.

More precisely, let $\mathcal{P}(x)$ be a property which, for $x \in X$, is satisfied or not by the sequence $(\Phi_n(x))$. If the set $\mathcal{A}_\mathcal{P} := \{x : \mathcal{P}(x) \text{ is true}\}$ is measurable and invariant by the map T, then by ergodicity of (X, μ, T) this set has measure 0 or 1: either $\mathcal{P}(x)$ is true for a.e. x, or $\mathcal{P}(x)$ is false for a.e. x.

Sometimes, for an asymptotic property $\mathcal{P}$, the set $\mathcal{A}_\mathcal{P}$ can be described in term of lim sup of a sequence of sets and its invariance by the map T can easily be checked.

The dichotomy between *recurrence and transience* of a cocycle is an example of a “0 -1” property: the property $\mathcal{R}$ “the cocycle is recurrent” corresponds to the set $\mathcal{A}_\mathcal{R} = \bigcup_{M\geq 1} \bigcap_{N\geq 1} \bigcup_{n\geq N} A_n^M$, where $A_n^M = \{x : \|\Phi_n(x)\| \leq M\}$.

Indeed, from the inequality (1.5) it follows $T^{-1}\mathcal{A}_\mathcal{R} = \mathcal{A}_\mathcal{R}$. Therefore, when (X, μ, T) is ergodic, either for μ a.e. x, $\lim_n \|\Phi_n(x)\| = +\infty$, or for a.e. x the cocycle $(\Phi_n(x))$ returns infinitely often in some compact set depending on x. In the latter case, an argument based on Poincaré recurrence lemma implies that the cocycle returns to any neighborhood of 0, for a.e. x.

We give below another example: the notion of limit direction.

2.1 Limit directions

Essential values and regularity

First we recall the classical notion of essential values of a recurrent cocycle with values in an abelian lcsc group G (cf. K. Schmidt [12]). A point ∞ is added to G with the natural notion of neighborhood. For our purpose, we restrict ourselves to the case $G = \mathbb{R}^d$.

Definition 2.1. An element $a \in G \cup \{\infty\}$ is *an essential value* of the cocycle (Φ, T) (with respect to μ) if, for every neighborhood $V(a)$ of a, for every measurable subset B of positive measure,

$$\mu\left(B \cap T^{-n}B \cap \{x : \Phi_n(x) \in V(a)\}\right) > 0, \text{ for some } n \geq 0. \tag{2.1}$$

The property (2.1) can be stated in the equivalent way:

$$\mu\left(\left\{x \in B : \Phi_n^B(x) \in V(a)\right\}\right) > 0, \text{ for some } n \geq 0. \tag{2.2}$$

We denote by $\overline{\mathcal{E}}(\Phi)$ the set of essential values of the cocycle (Φ, T) and by $\mathcal{E}(\Phi) = \overline{\mathcal{E}}(\Phi) \cap G$ the *set of finite essential values*.

Let us recall some facts. The set $\mathcal{E}(\Phi)$ is a closed subgroup of G. A cocycle Φ is a coboundary if and only if $\overline{\mathcal{E}}(\Phi) = \{0\}$. We have $\mathcal{E}(\Phi \bmod \mathcal{E}(\Phi)) = \{0\}$. Two cohomologous cocycles have the same set of essential values.

It is well known ([12], [1]) that the set $\mathcal{E}(\Phi)$ coincides with $\mathcal{P}(\Phi)$, the group of periods p of the measurable T_Φ-invariant functions on $X \times G$, i.e., the elements $p \in G$ such that for every T_Φ-invariant F, $F(x, y + p) = F(x, y), \lambda - a.e.$ This shows that $\mathcal{E}(\Phi) = G$ if and only if $(X \times G, \lambda, T_\Phi)$ is ergodic.

Definition 2.2. We say that the cocycle defined by Φ is *regular*, if it is cohomologous to a cocycle which has values in a closed subgroup H of G and is ergodic on $X \times H$. The group H in the definition is $\mathcal{E}(\Phi)$.

Now we consider the notion of limit directions and essential limit directions. The cocycle can be recurrent or transient.

Limit directions

For $v \in \mathbb{R}^d \setminus \{0\}$, let $\tilde{v} := v/\|v\|$ be the corresponding unit vector in the unit sphere $\mathbb{S}^{d-1}$. For every $\mathbb{R}^d$-valued cocycle (Φ_n), we obtain a process (directional process) $(\tilde{\Phi}_n)_{n\geq 1}$ with values in $\mathbb{S}^{d-1}$ (defined outside the values (n, x) such that $\Phi_n(x) = 0$).

Definition 2.3. A vector u is *a limit direction* of the cocycle $(\Phi_n(x))$ at x, if there exists a subsequence $(n_k(x))$ such that $\|\Phi_{n_k}(x)\| \to \infty$ and $\Phi_{n_k}(x)/\|\Phi_{n_k}(x)\|$ converges to u.

The subset for which the property $\mathcal{P}_u$: "u is a limit direction of $(\Phi_n(x))$" holds is

$$\mathcal{A}(u) = \bigcap_{V,M} \bigcap_{N} \bigcup_{n \geq N} \{x \in X : \left\|\Phi_n(x)\right\| > M \text{ and } \Phi_n(x)/\left\|\Phi_n(x)\right\| \in V\}. \tag{2.3}$$

where the intersection is taken over a countable basis of neighborhoods V of u and the positive integers M.

The set of limit directions of the cocycle $(\Phi_n(x))$ for $x \in X$ is defined as

$$\mathcal{D}(\Phi)(x) := \{u : \exists\, (n_k(x)) : \left\|\Phi_{n_k(x)}(x)\right\| \to \infty \text{ and } \Phi_{n_k(x)}(x)/\left\|\Phi_{n_k(x)}(x)\right\| \to u\}.$$

Remarks 1. a) From (1.5) it follows that $\mathcal{A}(u)$ is invariant by the map T, so that $\mathcal{P}_u$ is a "0 -1" property.

b) If Φ is integrable and $\int \Phi \, d\mu \neq 0$, then by the ergodic theorem $\mathcal{D}(\Phi)$ reduces to the direction defined by the mean of Φ. Therefore, when Φ is integrable, the interesting case is when $\int \Phi \, d\mu = 0$.

c) If Φ is a coboundary, $\Phi = \Psi - \Psi \circ T$, then the set of limit directions can be deduced from the support of the law of Ψ in $\mathbb{R}^d$. When this law gives a positive measure to each cone truncated from the origin, then, by ergodicity of T, the set $\mathcal{D}(\Phi)$ coincide with $\mathbb{S}^{d-1}$.

d) Billiards in the plane with periodic obstacles yield geometric examples of centered vector cocycles with a geometric interpretation of the limit directions (for these models, see for example [11], [13] for the dispersive billiards, [8], [4] for the billiards with polygonal obstacles).

Lemma 2.4. *There is a closed set $\mathcal{D}(\Phi)$ such that $\mathcal{D}(\Phi)(x) = \mathcal{D}(\Phi)$, for a.e. x. It is empty if and only if Φ is a coboundary: $\Phi = \Psi - \Psi \circ T$, with Ψ bounded.*

Proof. Clearly $\mathcal{D}(\Phi)(x)$ is a closed subset of $\mathbb{S}^{d-1}$. The invariance $\mathcal{D}(\Phi)(Tx) = \mathcal{D}(\Phi)(x)$ follows from (1.5). Using the Hausdorff distance on the set of closed subsets of $\mathbb{S}^{d-1}$ and ergodicity, we obtain that $\mathcal{D}(\Phi)(x)$ is a.e. equal to a closed subset $\mathcal{D}(\Phi)$.
If $\mathcal{D}(\Phi)$ is empty, then, for a.e. x, the sequence $(\Phi_n(x))$ is bounded. This implies that there is a measurable function Ψ such that $\Phi = \Psi - \Psi \circ T$. By ergodicity of T, Ψ is bounded. The converse is clear. $\square$

Definition 2.5. $\mathcal{D}(\Phi)$ will be called *set of limit directions* of (Φ_n). We write also $\mathcal{D}(T, \Phi)$ instead of $\mathcal{D}(\Phi)$ to explicit the dependence on T.

In other words, the "limit set" $\mathcal{D}(\Phi)$ is in the transient case the attractor in the sphere $\mathbb{S}^{d-1}$ of the process $(\tilde{\Phi}_n)_{n \geq 1}$ introduced above.

In the last section, we will show that under a strong stochastic hypothesis, this process $(\tilde{\Phi}_n)_{n \geq 1}$ visits any non empty open set in $\mathbb{S}^{d-1}$ and stays there during longer and longer intervals of time. This property can be formalized as follows.

Let (Z_n) be a process defined on (X, μ) with values in a metric space Y. A first question is about transitivity: does (Z_n) visit every non empty open set in Y. With the previous notion of limit direction, for the process $(\varphi_n(.)/\|\varphi_n(.)\|)$ associated to a transient cocycle (φ_n), this means $\mathcal{D}(\varphi) = \mathbb{S}^{d-1}$.

A stronger quantitative property is the following:

$$\limsup_n \frac{1}{n} \sum_1^n \mathbf{1}_V(Z_k(x)) = 1 \text{ a.e., for every non empty open set } V \text{ in } Y. \tag{2.4}$$

Clearly this property implies $\liminf_n n^{-1} \sum_{k=1}^n \mathbf{1}_V(Z_k(x)) = 0$ a.e., for every non empty open subset V in Y with a complement with non empty interior.

We will now discuss some general properties of the set of limit directions. The set of limit directions $\mathcal{D}(\Phi_B, T^B)$ for the induced cocycle Φ^B and the induced map T^B is denoted by $\mathcal{D}(\Phi_B)$ or $\mathcal{D}(B)$.

We have the equivalence:

Lemma 2.6. *a) A cocycle is a coboundary if and only if there is B of positive measure such that the set of limit directions for the induced cocycle on B is empty.*

b) If Φ and Φ' are cohomologous, there is B such that the corresponding induced cocycles on B have the same set of limit directions.

Proof. a) If u is a limit direction for (T_A, Φ_A), then it is also a limit direction for (T, Φ); hence the inclusion $\mathcal{D}(T_A, \Phi^A) \subset \mathcal{D}(T, \Phi)$.
By a compactness argument on the set of directions, if $(A_n)_{n\geq 1}$ is a sequence of decreasing sets with positive measure in X, then $(\mathcal{D}(T_{A_n}, \Phi^{A_n}))_{n\geq 1}$ is decreasing and the intersection is non empty, except if $\mathcal{D}(T_{A_{n_0}}, \Phi^{A_{n_0}})$ is empty for some n_0.

If (Φ_n) is bounded, or equivalently if $\Phi = \Psi - \Psi \circ T$, with a bounded Ψ, then clearly $\mathcal{D}(T, \Phi)$ is empty.

Suppose now that Φ is a coboundary, $\Phi = \Psi - \Psi \circ T$, with Ψ measurable. Let B be such a set such that Ψ is bounded on B. Then the induced cocycle $(\Phi_n^B)_{n\geq 1}$ is bounded, since $\Phi_n^B = \Psi - \Psi \circ T_B^n$. Therefore $\mathcal{D}(T_B, \Phi^B)$ is empty.

Conversely, if there is B of positive measure in X such that $\mathcal{D}(T_B, \Phi^B)$ is empty, then the induced cocycle (Φ_n^B) is bounded, so Φ^B is a T_B cocycle. By Lemma 2.7 below, Φ is a coboundary.

b) Let Φ and Φ' be such that $\Phi' = T\Psi - \Psi + \Phi$ for a measurable Ψ. Let B such that Ψ is bounded on B. Then $\Phi'^B_n = T_B^n\Psi - \Psi + \Phi_n^B$ with $T_B^n\Psi - \Psi$ bounded, which implies that (Φ'^B_n) and (Φ_n^B) have the same limit directions. □

Lemma 2.7. *Let B be such that $X = \cup_{k\geq 0} T^k B$. If Φ^B is a T_B-coboundary, then Φ is a T-coboundary.*

Proof. For μ-a.e. $y \in X$ there are a unique $x \in B$ and an integer k, $0 \leq k < R_B(x)$, such that $y = T^k x$. Suppose that there is Ψ on B such that: $\Phi^B = \Psi - \Psi \circ T^B$. We define ζ on X by taking, for $0 \leq k < R_B(x)$, $\zeta(T^k x) = \Psi(x) - \Phi_k(x)$.

We have $\Phi(y) = \zeta(y) - \zeta(Ty)$. Indeed, for $y = T^k x$, $0 \leq k < R_B(x) - 1$, the relation is satisfied by construction; for $y = T^k x$ with $k = R_B(x) - 1$, the relation follows from the coboundary relation for the induced cocycle. □

Now let us show that two sets A_1 and A_2 have always non disjoint sets of limit directions, unless Φ is a coboundary.

Lemma 2.8. *For any two sets A_1 and A_2, there is $B_1 \subset A_1$ such that $\mathcal{D}(B_1) \subset \mathcal{D}(A_2)$.*

Proof. Let $B_1 \subset A_1$ be such that Φ^{A_2} (as defined by (1.3)) satisfies $\Phi^{A_2}(x) \leq C$ on B_1, for some constant C. If $\mathcal{D}(B_1)$ is empty, then Φ is a coboundary by Lemma 2.6. Let u be a limit direction for Φ^{B_1}.

The cocycle $(\Phi_n^{A_2}(x))$, for a piece of orbit starting and ending in A_2 and for a special sequence of times, can be written as (1) + (2) + (3) where

$$\begin{aligned}
(1) &= \sum_{t=0}^{R_{B_1}(x)-1} \Phi\left(T^t x\right), \\
(2) &= \Phi^{B_1}_{n_k(x_1)}(x_1), \text{ with } x_1 = T^{R_{B_1}(x)}x, \\
(3) &= \Phi^{A_2}(x_2), \text{ with } x_2 = T_{B_1}^{n_k(x_1)}x_1.
\end{aligned}$$

The first term (1) corresponds to the path from A_2 to B_1. The second term (2) corresponds to visits of the cocycle induced on B_1 to a neighborhood of u (after normalization) with an arbitrary large norm (such visits exist because u is a limit direction for the induced cocycle on B_1), and the third (3) to the path from B_1 to A_2 with a bounded value of the cocycle by construction.

If we iterate for a long time the induced cocycle (2), the first term (which is fixed) and the third (which is bounded) are small compared with the norm of (2). Then (1) + (2) + (3) gives a value of the induced cocycle on A_2 which satisfy the condition that u is a limit direction for Φ^{A_2}. □

2.2 Essential limit directions

The observation that the set of limit directions $\mathcal{D}(\Phi)$ is not a "cohomological invariant" motivates the following definition.

Definition 2.9. A direction $u \in \mathbb{S}^{d-1}$ is called an *essential limit direction* for Φ, if, for every subset B of positive measure, u is a limit direction for Φ^B. The set of essential limit directions is denoted by $\mathcal{ED}(\Phi)$.

The set $\mathcal{ED}(\Phi)$ can be seen as a "boundary" for (Φ_n). It is invariant by cohomology: if Φ_1 and Φ_2 are cohomologous, then $\mathcal{ED}(\Phi_1) = \mathcal{ED}(\Phi_2)$.

Theorem 2.10. *1) $\mathcal{ED}(\Phi)$ is a closed subset of $\mathbb{S}^{d-1}$ which is empty if and only if Φ is a coboundary. For every B of positive measure, $\mathcal{ED}(\Phi^B) = \mathcal{ED}(\Phi)$.*

2) If (Φ_n) is transient and Φ is bounded, then $\mathcal{D}(\Phi)$ is a closed connected non empty subset of $\mathbb{S}^{d-1}$.

Proof. 1) We have $\mathcal{ED}(\Phi) = \bigcap \mathcal{D}(\Phi^B)$, where the intersection is over the family of all measurable subsets of positive measure.

Clearly, $\mathcal{ED}(\Phi) \subset \mathcal{ED}(\Phi^B)$. Let A with $\mu(A) > 0$. By Lemma 2.8, there is $B_1 \subset B$ such that every limit direction for Φ^{B_1} is a limit direction for Φ^A. If u is in $\mathcal{ED}(\Phi^B)$, then u

is in $\mathcal{D}(\Phi^{B_1})$, hence in $\mathcal{D}(\Phi^A)$. Therefore $\mathcal{ED}(\Phi^B) \subset \mathcal{D}(\Phi^A)$, for all A, which implies $\mathcal{ED}(\Phi^B) \subset \mathcal{ED}(\Phi)$.

2) Let u_1 and u_2 be two accumulation points of $\Phi_n(x)/\|\Phi_n(x)\|$ and $\varepsilon > 0$. For a.e. x, by transience, for N big enough, we have $\|\Phi_n(x)\| \geq \varepsilon^{-1}\|\Phi\|_\infty, \forall n \geq N$. By definition, there exist $n > m > N$ such that $d(\Phi_m(x)/\|\Phi_m(x)\|, u_1) < \varepsilon$ and $d(\Phi_n(x)/\|\Phi_n(x)\|, u_2) < \varepsilon$. Moreover, for every k between m and $n-1$, one has

$$\begin{aligned}
&\left\| \frac{\Phi_k(x)}{\|\Phi_k(x)\|} - \frac{\Phi_{k+1}(x)}{\|\Phi_{k+1}(x)\|} \right\| \\
&\leq \frac{\|\Phi \circ T^k(x)\|}{\|\Phi_{k+1}(x)\|} + \|\Phi_k(x)\| \left| \frac{1}{\|\Phi_{k+1}(x)\|} - \frac{1}{\|\Phi_k(x)\|} \right| \leq 2\varepsilon.
\end{aligned}$$

Thus, for every $\varepsilon > 0$, one has a finite set F_ε of points $\Phi_k(x)/\|\Phi_k(x)\|$ on the unit sphere that can be used to go from u_1 to u_2 with jumps of length smaller than 2ε. Now let ε tend to zero and consider F_∞ an accumulation point of $(F_\varepsilon)_{\varepsilon>0}$ in the set of compact sets of the sphere equipped with the Hausdorff metric. The set F_∞ is a connected compact set containing u_1 and u_2. □

Let $\widetilde{\mathcal{E}}(\Phi)$ be the smallest vector space of $\mathbb{R}^d$ containing $\mathcal{E}(\Phi)$. Using Definition 2.2, we have:

Theorem 2.11. *For every non coboundary Φ, $\mathcal{ED}(\Phi)$ contains $\mathbb{S}(\widetilde{\mathcal{E}}(\Phi))$, the sphere at infinity of $\widetilde{\mathcal{E}}(\Phi)$, and is equal to $\mathbb{S}(\widetilde{\mathcal{E}}(\Phi))$ if Φ is a regular cocycle.*

Remarks and questions. a) A general question is to find the set of limit directions and the set of essential limit directions of a given cocycle. What are the possible shapes of these sets ?
b) The rate of growth of the cocycle plays no role in the "directional process" associated to a cocycle as defined above. This rate could be taken into account by introducing a scaling in the notion of limit directions.

c) Let us call "irreducible" a $\mathbb{R}^d$-cocycle which is not cohomologous to a cocycle with values in a vector subspace of dimension $< d$. For an irreducible cocycle Φ what kind of set $\mathcal{D}(\Phi)$ can be? In particular does there exist a recurrent cocycle (Φ_n) such that $\mathcal{D}(\Phi)$ reduces to two antipodal points?

These questions are related to the following remark. Let Φ and Ψ with values in $\mathbb{R}^d$ be given. We say that *the cocycle (Φ_n) dominates (Ψ_n), if there are C and K such that $\|\Psi_n(x)\| \leq C\|\Phi_n(x)\| + K, \forall n$.*

Clearly this is the case when Ψ is cohomologous to a multiple of Φ with a bounded transfer function. The proposition below is a partial converse.

Proposition 2.12. *Assume that T_Φ is ergodic on $X \times \mathbb{R}^d$. If (Φ_n) dominates (Ψ_n), then Ψ is cohomologous to $c\,\Phi$ for a constant c.*

Proof. Let I be a compact neighborhood of $\{0\}$. Let y be in I. For the times $n_k(x,y)$ such that $y + \Phi_{n_k(x,y)}(x) \in I$, $|\Psi_n(x)|$ is bounded. The cocycle $\Psi_n^{Z_I}(x)$ induced of Ψ

on the set $Z_I := X \times I$ is bounded. Therefore the function F defined on $X \times \mathbb{R}$ by $F(x,y) = \Psi(x)$ is a coboundary for the map T_Φ: there is $H(x,y)$ such that $F(x,y) = \Psi(x) = H(Tx, y + \Phi(x)) - H(x,y)$.
For every $a \in \mathbb{R}^d$, the function $(x,y) \to H(x,y+a)-H(x,y)$ is T_Φ-invariant, hence a.e. constant by ergodicity of T_Φ: for a.e. (x,y), there is $c(a)$ such that $H(x,y+a) = c(a) + H(x,y)$. By the theorem of Fubini, for a.e. y, $H(x,y+a) = c(a)+H(x,y)$, for almost every (x,a), and $a \to c(a)$ is Lebesgue measurable. Let us take y_0 satisfying this property. We have $H(x, a+y_0) = c(a) + H(x,y_0)$, for a.e. (x,a); hence, with $u(a) = c(a-y_0)$ and $h(x) = H(x,y_0)$:

$$H(x,a) = u(a) + h(x), \text{ for a.e. } (x,a).$$

The relation $H(x,y+a) = c(a)+H(x,y)$ reads: $u(y+a)+h(x) = c(a)+u(y)+h(x)$ which shows that u is an affine function.

Therefore $H(x,y) = c_1 y + c_2 + h(x)$ for constants c_1, c_2 and a measurable function h on X and we have $\Psi(x) = c\,\Phi(x) + h(Tx) - h(x)$. □

2.3 A G_δ-property

Suppose that the map $T = T(\theta)$ and the function defining the cocycle $\Phi = \Phi^\theta$ depend on a parameter θ. Suppose that Θ, the set of parameters, is a metric space and that the dependence of $T(\theta)$ and Φ^θ is continuous. In the following theorem we take for X a compact metric space and the measure is supposed to be regular on X. (Remark that the result can be extended to the piecewise continuous case). We denote by $\mathcal{V} = \mathcal{V}(u)$ a countable basis of open neighborhoods of a direction u in $\mathbb{S}^{d-1}$.

Theorem 2.13. *Suppose that in the set of parameters Θ there is a dense set $\mathcal{T}$ of values such that the corresponding set of limit directions is $\mathbb{S}^{d-1}$. Then there is a dense G_δ-set in Θ with the same property.*

Proof. We can assume that $\mathcal{T}$ is countable: $\mathcal{T} = \{\theta_i, i = 1,2,...\}$. Fix a direction $u \in \mathbb{S}^{d-1}$. For $\theta \in \mathcal{T}$, for a.e. $x \in X$, u is a limit direction for $(\Phi_n^\theta(x))$. Let K be a compact set of positive measure in X such that for every i,

$$K \subset \left\{x \in X : u \text{ is a limit direction for } \Phi_n^{\theta_i}(x)\right\}.$$

For a fixed x, for $M \geq 1$ and $V \in \mathcal{V}(u)$, the set

$$\tilde{B}_n^{x,V,M} = \left\{\theta : \left\|\Phi_n^\theta(x)\right\| > M \text{ and } \Phi_n^\theta(x)/\left\|\Phi_n^\theta(x)\right\| \in V\right\}$$

is an open set.

If W is an open set in X, let

$$\tilde{B}_n^{W,V,M} := \left\{\theta : \left\|\Phi_n^\theta(y)\right\| > M \text{ and } \Phi_n^\theta(y)/\left\|\Phi_n^\theta(y)\right\| \in V, \forall y \in W\right\}.$$

Let $V \in \mathcal{V}(u)$, $M \in \mathbb{N}$ and $\theta_i \in \mathcal{T}$. For each $x \in K$, there exists n such that $\theta_i \in \tilde{B}_n^{x,V,M}$. By continuity $\theta_i \in \tilde{B}_n^{y,V,M}$ for y in an open neighborhood of x. Thus there are finitely many open sets $W^1_{(V,M)}, \dots, W^{r_i(V,M)}_{(V,M)}$ covering K and integers $n^1_{(V,M)}, \dots, n^{r_i(V,M)}_{(V,M)}$ such that

$$\theta_i \in \bigcap_{j=1,\dots,r_i(V,M)} \tilde{B}^{W^j_{(V,M)},V,M}_{n^j_{(V,M)}}.$$

This proves that, for every $y \in K$, there are $j \in \{1, \dots, r_i(V, M)\}$ and n_j such that

$$\left\|\Phi^{\theta_i}_{n_j}(y)\right\| > M \text{ and } \Phi^{\theta_i}_{n_j}(y)/\left\|\Phi^{\theta_i}_{n_j}(y)\right\| \in V.$$

For every i, θ_i belongs to the open set

$$\bigcup_i \bigcap_{j\in\{1,\dots,r_i\}} \tilde{B}^{W^j_{(V,M)},V,M}_{n^j_{(V,M)}}.$$

The dense set of parameters $\mathcal{T} = \{\theta_i\}$ is contained in the countable intersection of open sets:

$$\bigcap_{V\in\mathcal{V}(u),M\geq 1} \bigcup_i \bigcap_{j\in\{1,\dots,r_i(V,M)\}} \tilde{B}^{W^j_{(V,M)},V,M}_{n^j_{(V,M)}}. \tag{2.5}$$

Now, suppose that the parameter belongs to the dense G_δ-set defined above by (2.5).

For every $V \in \mathcal{V}(u)$ and $M \in \mathbb{N}$, there is i such that

$$\theta \in \bigcap_{j\in\{1,\dots,r_i(V,M)\}} \tilde{B}^{W^j_{(V,M)},V,M}_{n^j_{(V,M)}},$$

i.e., for every $V \in \mathcal{V}(u)$ and $M \in \mathbb{N}$, there are i and $j \in \{1, \dots, r_i(V, M)\}$ such that

$$\forall y \in W^j_{(V,M)}, \left\|\Phi^{\theta}_{n^j_{(V,M)}}(y)\right\| > M \text{ and } \Phi^{\theta}_{n^j_{(V,M)}}(y)/\left\|\Phi^{\theta}_{n^j_{(V,M)}}(y)\right\| \in V.$$

As $W^j_{(V,M)}$, $j = 1, \dots, r_i(V, M)$, is a covering of K, for each point $y \in K$, for all $V \in \mathcal{V}(u)$, all $M \in \mathbb{N}$, there is n such that $\|\Phi^\theta_n(y)\| > M$ and $\Phi^\theta_n(y)/\|\Phi^\theta_n(y)\| \in V$.

Therefore for each $y \in K$, u is a limit direction. As it is a 0 -1-property, the property that u is a limit direction holds for a.e. x. □

2.4 Limit directions and limit distributions

Lemma 2.14. *Suppose that, for a sequence of integers (k_n) and a sequence (a_n) tending to ∞, (Φ_n) satisfies a limit theorem in distribution:*

$$a_n^{-1}\Phi_{k_n} \xrightarrow{distrib} \mathcal{L},$$

where $\mathcal{L}$ is a probability measure on $\mathbb{R}^d$ giving a positive probability to each non empty open set. Then the set $\mathcal{D}(\Phi)$ of limit directions of (Φ_n) is $\mathbb{S}^{d-1}$. This applies in particular if (Φ_n) satisfies a non degenerated CLT for a subsequence and an adapted normalization.

Proof. Suppose that $u \in \mathbb{S}^{d-1}$ is not a limit direction for (Φ_n). By the dichotomy (cf. Remark 1 a)), there is M and an open regular neighborhood $V = V(u)$ of u in $\mathbb{R}^d$ such that, a.e. x belongs for some N to the set

$$C_N := \left\{x : \forall n \geq N, \left\|\Phi_{k_n}(x)\right\| \leq M \text{ or } \Phi_{k_n}(x)/\left\|\Phi_{k_n}(x)\right\| \notin V(u)\right\}.$$

The sequence of sets (C_N) is increasing and $\mu(\bigcup_N C_N) = 1$.

From the assumption, we have

$$\liminf_{n\to+\infty} \mu\left\{x \in X : a_n^{-1}\Phi_{k_n}(x) \in V\right\} \geq \mathcal{L}(V) > 0.$$

Therefore for any $\alpha > 0$, there is N such that, for $n \geq N$, there is a set B in X of measure $> \frac{\mathcal{L}(V)}{2}$ such that: $\|\Phi_{k_n}(x) - a_n u\| \leq \alpha a_n$, which implies

$$(1-\alpha)a_n \leq \left\|\Phi_{k_n}(x)\right\| \leq (1+\alpha)a_n, \text{ for } x \in B.$$

Hence: $\|\Phi_{k_n}(x)/\|\Phi_{k_n}(x)\| - u\| \leq \frac{\alpha}{1-\alpha}$ on a set of measure $\frac{\mathcal{L}(V)}{2} > 0$. If we take N such that $\mu(C_N) > 1 - \frac{\mathcal{L}(V)}{2}$, there is a contradiction for $n > N$ big enough.

This applies when $\mathcal{L} = \mathcal{N}(0,\Gamma)$ where Γ is a non degenerated covariance matrix. □

2.5 Oscillations of 1-dimensional cocycles

We discuss now the notion of limit directions in the special case of cocycles with values in $\mathbb{R}$. About oscillations of 1-dimensional cocycles, let us mention the work of Derriennic ([5]) where other references on the subject, in particular of Tanny [14] and Woś ([15]), can also be found. For completeness we give below a short presentation, related to the notion of limit direction, of some results on 1-dimensional cocycles.

We consider a conservative transformation T of a space (X,μ) where μ is σ-finite and non singular for T. Notice that in Lemmas 2.15 and 2.16 below, ergodicity is not assumed. Recall that equalities between functions are understood μ-a.e.

In this subsection, we denote by φ and (φ_n) respectively a given measurable function and the corresponding cocycle over (X,μ,T). We define two sets, clearly T-invariant (cf. (1.5)):

$$F_\varphi^+ := \left(\inf_n \varphi_n(.) > -\infty\right),\ F_\varphi^- := \left(\sup_n \varphi_n(.) < +\infty\right). \tag{2.6}$$

Let us recall the following classical lemma (filling scheme).

Lemma 2.15. *There are two functions h^+ and g^+ defined on F_φ^+ (resp. h^- and g^- defined on F_φ^-) with values in $[0, +\infty[$ such that*

$$\begin{aligned} \varphi(x) &= h^+(Tx) - h^+(x) + g^+(x), \text{ for } \mu\text{–a.e. } x \in F_\varphi^+, &(2.7)\\ \varphi(x) &= -h^-(Tx) + h^-(x) - g^-(x), \text{ for } \mu\text{–a.e. } x \in F_\varphi^-, &(2.8)\end{aligned}$$

On the invariant set $F_\varphi^{+\infty} := (\varphi_n(.) \to +\infty)$, we have $\sum_{k=0}^\infty g^+(T^k x) = +\infty$.

If (φ_n) is recurrent, then the space X decomposes in two invariant sets, each of them possibly of zero measure, one on which φ is a coboundary, the other on which $\sup_n \varphi_n(.) = +\infty$ and $\inf_n \varphi_n(.) = -\infty$.

Proof. Let $m_n(x) := \min_{1\le k\le n}(\varphi_k(x))$, $n \ge 1$. We have

$$m_{n+1}(x) = \min(\varphi(x), \varphi(x) + m_n(Tx)) = \begin{cases} \varphi(x) - m_n^-(Tx), \text{if } m_n(Tx) \le 0,\\ \varphi(x) = \varphi(x) - m_n^-(Tx), \text{if } m_n(Tx) > 0,\end{cases}$$

which implies $m_{n+1}(x) = \varphi(x) - m_n^-(Tx)$. Since the limit $m_\infty(x) := \lim_n m_n(x)$ is finite on F_φ^+, it follows:

$$\varphi(x) = m_\infty^-(Tx) - m_\infty^-(x) + m_\infty^+(x),\ x \in F_\varphi^+.$$

This gives the decomposition (2.7) on F_φ^+, with $h^+ = m_\infty^-$ and $g^+ = m_\infty^+$. We get (2.8) by changing φ into $-\varphi$.

On the invariant set $F_\varphi^{+\infty} = (\varphi_n(.) \to +\infty)$ the decomposition (2.7) holds. Let us show that $\sum_{k=0}^\infty g^+(T^k x) = +\infty$.

We have $\varphi_n(x) = h^+(T^n x) - h^+(x) + \sum_{k=0}^{n-1} g^+(T^k x)$. Let M_K be the subset of $F_\varphi^{+\infty}$ where h^+ is bounded by a finite constant K. By conservativity of T, for a.e. x in M_K there is a subsequence $(n_j(x))$ such that $T^{n_j(x)}(x) \in M_K$ and therefore $\sum_{k=0}^{n_j(x)-1} g^+(T^k x) \to +\infty$. It follows $\sum_{k=0}^\infty g^+(T^k x) = +\infty$ for a.e x in M_K and, since K is arbitrary, $\sum_{k=0}^\infty g^+(T^k x) = +\infty$ a.e.

Suppose now that (φ_n) is recurrent. Let $g_\infty^+(x) = \sum_0^\infty g^+(T^k x) \in [0, +\infty]$. The induced cocycle (φ_n^B) is recurrent for any set B on which h^+ is bounded and therefore $g_\infty^+(x) < +\infty$, a.e. on B. As the sets B cover F_φ^+, this implies $g_\infty^+(x) < +\infty$, a.e. on F_φ^+. Therefore $g^+(x) = g_\infty^+(x) - g_\infty^+(Tx)$, and the restriction of φ to the invariant set F_φ^+ is a coboundary. Likewise, φ is a coboundary on the invariant set F_φ^-.

So we have proved that, for any recurrent cocycle, the space X decomposes in two sets, the set $F_\varphi^+ \cup F_\varphi^-$ on which φ is a coboundary and its complement on which φ_n oscillates between $+\infty$ and $-\infty$. □

The lemma implies that φ is a coboundary on the invariant set $\{x : \varphi_n(x) \text{ is bounded}\}$. It is well known that, if (φ_n) is uniformly bounded, then the transfer function is bounded.

The previous lemma gives a simple way to prove and to slightly extend a result of Kesten on the rate of divergence in dimension 1 of a non recurrent cocycle. We consider a conservative dynamical system with a σ-finite invariant measure μ. The σ-algebra of T-invariant sets is denoted by $\mathcal{I}$.

As μ is σ-finite, we can choose a function p on X such that $\mu(p) = 1$ and $0 < p(x) \leq 1, \forall x$. By the ratio ergodic theorem, for $f \in \mathbb{L}^1(\mu)$,

$$\lim_{n\to\infty} \frac{\sum_{k=0}^{n-1} f\left(T^k x\right)}{\sum_{k=0}^{n-1} p\left(T^k x\right)} = \mathbb{E}_{p\mu}\left[\frac{f}{p} \,|\mathcal{I}\right](x),\ \mu\text{-a.e. } x \in X.$$

We have $p_n(x) = \sum_0^{n-1} p(T^k x) \to \infty$ since the system is conservative.

Lemma 2.16. *(cf. also [9], [5]) Suppose that the T-invariant measure μ is conservative σ-finite.*

1) If φ is a non negative measurable function, then for μ-a.e. x the sum $\sum_0^\infty \varphi(T^k x)$ is either 0 or $+\infty$, and $\liminf_n \varphi_n(x)/p_n(x) \in]0, +\infty]$ on the set $\{\sum_0^\infty \varphi(T^k.) \neq 0\}$.

2) For any measurable φ, $\liminf_n \varphi_n(x)/p_n(x) > 0$ for μ-a.e. x on the set $(\varphi_n(.) \to +\infty)$.

The cocycle (φ_n) is recurrent on the invariant set $\{x : \lim_n \varphi_n(x)/p_n(x) = 0\}$.

3) If μ is a T-invariant probability measure, then, for any measurable φ, $\liminf_n \frac{1}{n}\varphi_n(x) > 0$ for μ-a.e. x, on the set $(\varphi_n(.) \to +\infty)$ and the cocycle (φ_n) is recurrent on the invariant set $\{x : \lim_n \frac{1}{n}\varphi_n(x) = 0\}$.

Proof. 1a) First, suppose that $\varphi = 1_B$ where B is a measurable subset. For $N \geq 0$, let $B^{(N)} := \cup_0^N T^{-k}B$, $B^{(\infty)} = \cup_0^\infty T^{-k}B$. We have $T^{-1}B^{(\infty)} \subset B^{(\infty)}$, hence $T^{-1}B^{(\infty)} = B^{(\infty)}$ up to a negligible set as (X, μ, T) is conservative.

On the complement of $B^{(\infty)}$, we have $\sum_0^n 1_B(T^k x) = 0$. By the ratio ergodic theorem

$$\lim_n \frac{1}{p_n(x)}\ \left(p_n(x)\right)^{-1} \sum_0^{n-1} 1_B\left(T^k x\right)\, p\left(T^k x\right) = \mathbb{E}_{p\mu}\left(1_B \middle| \mathcal{I}\right)(x).$$

As $\sum_{j=0}^{L-1} 1_B(T^j x) \geq 1_{B^{(L)}}(x)$ and $\lim_n p_n(x) = +\infty$, we have

$$\begin{aligned}\liminf_n \frac{1}{p_n(x)} \sum_{k=0}^{n-1} 1_B\left(T^k x\right) &\geq \frac{1}{L} \liminf_n \frac{1}{p_n(x)} \sum_{k=0}^{n-1} 1_{B^{(L)}}\left(T^k x\right) \\ &\geq \frac{1}{L} \lim_n \frac{1}{p_n(x)} \sum_{k=0}^{n-1} 1_{B^{(L)}}\left(T^k x\right) p\left(T^k x\right) = \frac{1}{L}\mathbb{E}_{p\mu}\left(1_{B^{(L)}} \middle| \mathcal{I}\right)(x).\end{aligned}$$

Therefore from the relation

$$\bigcup_L \uparrow \left\{x : \mathbb{E}_{p\mu}\left(1_{B^{(L)}} \middle| \mathcal{I}\right)(x) > 0\right\} = \left\{x : \mathbb{E}_{p\mu}\left(1_{B^{(\infty)}} \middle| \mathcal{I}\right)(x) > 0\right\} = 1_{B^{(\infty)}},$$

it follows

$$\liminf_n \frac{1}{p_n(x)} \sum_{k=0}^{n-1} 1_B\left(T^k x\right) > 0 \text{ on } B^{(\infty)}. \tag{2.9}$$

1b) Now let φ be any non negative function. For $j \in \mathbb{Z}$, let $B_j := \{2^j \leq \varphi < 2^{j+1}\}$. We get: $\sum_0^\infty \varphi(T^k x) > 0 \Leftrightarrow \exists j : x \in B_j^\infty$. Using (2.9) applied to the sets B_j and the inequality $\varphi \geq \sum_{j=-\infty}^{+\infty} 2^j 1_{B_j}$, we obtain:

$$\begin{aligned}&\liminf_n \frac{1}{p_n(x)} \sum_0^{n-1} \varphi\left(T^k x\right) \\ &\geq \sum_{j=-\infty}^{+\infty} 2^j \liminf_n \frac{1}{p_n(x)} \sum_0^{n-1} 1_{B_j}\left(T^k x\right) > 0, \text{ on } \left(\sum_0^\infty \varphi\left(T^k .\right) > 0\right).\end{aligned}$$

2) As (2.7) of Lemma 2.15 holds on the set $F_\varphi^{+\infty} = (\varphi_n(.) \to +\infty)$ and $\sum_{k=0}^\infty g^+ (T^k x) = +\infty$, we can apply 1) to g^+. Since $\varphi_n(x) \geq -h^+(x) + g_n^+(x)$, we get:

$$\liminf_n \frac{\varphi_n(x)}{p_n(x)} \geq \liminf_n \frac{g_n^+(x)}{p_n(x)} > 0, \ \text{ for } \mu\text{−a.e. } x \in F_\varphi^{+\infty}.$$

This implies that $(\varphi_n(.))$ is recurrent on the invariant set $\{x : \lim_n \varphi_n(x)/p_n(x) = 0\}$, since we can not have $\varphi_n(x) \to +\infty$ or $\varphi_n(x) \to -\infty$ on this set.

In particular, if T is ergodic and φ integrable with $\mu(\varphi) = 0$, we have $\lim_n \varphi_n(x)/p_n(x) = \mu(\varphi) = 0$, which implies the recurrence of the cocycle (φ_n).

3) When the measure is finite, then we take $p(x) = 1$ and $p_n(x)$ is replaced by n. □

Proposition 2.17. *For a 1-dimensional cocycle (φ_n) generated by φ over an ergodic dynamical system, if φ is not a coboundary, one of the following (exclusive) properties is satisfied:*
1) $\limsup_n \varphi_n = -\liminf \varphi_n = +\infty$,
2) φ_n *tends to* $+\infty$,
3) φ_n *tends to* $-\infty$.

Proof. By ergodicity, with the notation (2.6), one of the sets $(F_\varphi^+ \cup F_\varphi^-)^c$, F_φ^+, F_φ^- has full measure.
The first case is equivalent to property 1) and to the equality $\mathcal{ED}(\varphi) = \{-\infty, +\infty\}$. Suppose now that F_φ^+ has full measure. Then we have the decomposition (2.7), Lemma 2.15, with equality a.e. Hence $\varphi_n(x) \geq -h^+(x) + g_n^+(x)$, for μ−a.e. x. Since φ is not a coboundary, g^+ is non negligible. This implies property 2). Likewise 3) holds if F_φ^- has full measure. □

This leads to the following remarks.
- if the cocycle (φ_n) is recurrent, it oscillates between $+\infty$ and $-\infty$, unless φ is a coboundary with a transfer function bounded from above or below;
- if φ is a coboundary, $\varphi = T\psi - \psi$, then if ψ is not essentially bounded from above (resp. from below), then $\limsup_n \varphi_n = +\infty$ (resp. $\liminf \varphi_n = -\infty$);
- we have $-\infty \notin \mathcal{D}(\varphi)$ if and only if $\varphi = Th - h + g$, with h, g non negative and g non negligible. This is equivalent to $\lim_n \varphi_n = +\infty$;
- $\mathcal{D}(\varphi)$ is empty if φ is a coboundary, $\varphi = Th - h$, with h essentially bounded.

For the constructions below, we need the following lemma:

Lemma 2.18. *Let (ℓ_n) be a strictly increasing sequence of integers. For any ergodic dynamical system there is h non negative such that, for a.e. x, $h(T^n x) \geq \ell_n$ infinitely often.*

Proof. There exists a strictly increasing sequence of positive real numbers (c_j) and a strictly increasing sequence of natural numbers (n_j), both tending to infinity, and a non negative measurable function f such that

$$\lim_j \frac{1}{c_j n_j} \sum_{k=1}^{n_j} f\left(T^k x\right) = +\infty, \text{ a.e.}$$

We put $d_k = c_j$, for $n_{j-1} \leq k < n_j$. For a.e. x, for j big enough, we can define a non decreasing sequence $(k_j(x))$ such that $\lim_j k_j(x) = +\infty$ and $f(T^{k_j(x)}x) \geq c_j \geq d_{k_j(x)}$. (Put $k_j(x) := \max\{k \leq n_j : f(T^k x) \geq c_j\}$.)
Now we can define a non decreasing function γ on $\mathbb{R}^+$ by putting $\gamma(y) = \ell_k$, for $d_k \leq y < d_{k+1}$. In particular, we have $\gamma(d_k) = \ell_k$.

Let $h(x) := \gamma(f(x))$. Then we have: $h(T^k x) = \gamma(f(T^k x)) \geq \gamma(d_k) = \ell_k$, if $f(T^k x) \geq d_k$.

Therefore for a.e. x, for j big enough, $h(T^{k_j(x)}(x)) \geq \ell_{k_j(x)}$. □

For every $B \subset X$ of positive measure, φ is a coboundary, if and only if the induced cocycle φ^B is a coboundary for the induced map T_B. If φ is not a coboundary, then the inclusion $\mathcal{D}(\varphi^B) \subset \mathcal{D}(\varphi)$ is general, but can be strict.

Example: Let B with $\mu(B) > 0$. Take $\varphi = Th - h + 1$, with $h(x) \geq 0$ on B. Then, if $R_n(x)$ denotes the n-th return time to B, we have

$$\varphi_n^B(x) = h\left(T_B^n x\right) - h(x) + R_n(x) \geq -h(x) + R_n(x) \to +\infty.$$

We can choose h on B^c such that, for the cocycle $\varphi_n(x) = h(T^n x) - h(x) + n$, we have $h(T^n x) \leq -n^2$ infinitely often (Lemma 2.18). Therefore $-\infty \in \mathcal{D}(\varphi) \neq \mathcal{D}(\varphi_B) = \{+\infty\}$.

Reverse cocycle
Recall that the reverse cocycle $(\check{\varphi}_n)_{n\geq 0}$ is defined by $\check{\varphi}_0 = 0$ and

$$\check{\varphi}_n(x) = -\varphi_n\left(T^{-n}x\right) = -\sum_{k=1}^{n} \varphi\left(T^{-k}x\right), \text{ for } n \geq 1.$$

For an ergodic system, if φ is integrable and if $\lim_n \varphi_n = +\infty$, then $\lim_n \check{\varphi}_n = -\infty$, since both conditions are equivalent to $\mu(\varphi) > 0$.

If φ is non integrable, we can have $\lim_n \varphi_n = +\infty$ and $\limsup_n \check{\varphi}_n = +\infty$. (see also [5]).

Example: Let $\varphi = Th - h + 1$, with h non negative. We have $\lim_n \varphi_n(x) = +\infty$. The reverse cocycle reads

$$\check{\varphi}_n(x) = -\varphi_n\left(T^{-n}x\right) = -h(x) + h\left(T^{-n}x\right) - n.$$

If h is chosen such that the inequality $h(T^{-n}x) \geq n^2$ occurs infinitely often for a.e. x (Lemma 2.18), then $+\infty$ is a limit direction for the reverse cocycle.

3.7 A result in dimension 2

Let us mention a partial result for 2-dimensional cocycles:

Proposition 2.19. *Let $\Phi : X \to \mathbb{R}^2$ be an integrable and centered function. If (Φ_n) is a transient cocycle over an ergodic dynamical system, then $\mathcal{D}(\Phi) \cup (-\mathcal{D}(\Phi)) = \mathbb{S}^1$ for a.e. $x \in X$.*

Proof. We denote by $< u, v >$ the scalar product in $\mathbb{R}^2$. Let $v \in \mathbb{S}^1$, and let $v^\perp \in \mathbb{S}^1$ be such that $\langle v, v^\perp \rangle = 0$. The function $x \mapsto \langle \Phi(x), v^\perp \rangle$ has zero integral and T is ergodic, hence the cocycle $\langle (\Phi_n), v^\perp \rangle$ is recurrent. For a.e. $x \in X$ there is a sequence $n_k(x) \to \infty$ and $c > 0$ such that $|\langle \Phi_{n_k}(x), v^\perp \rangle| < c$. As (Φ_n) is transient, for a.e. x $|\langle \Phi_{n_k}(x), v \rangle|$ is not bounded. There is a subsequence (n_{k_j}) such that $\Phi_{n_{k_j}}(x)/\|\Phi_{n_{k_j}}(x)\|$ converges to v or $-v$, i.e., v or $-v \in \mathcal{D}(\Phi)(x)$. □

By what precedes and Theorem 2.10, when Φ is bounded, the set $\mathcal{D}(\Phi)$ is an arc of a circle with length $\geq \frac{1}{2}$.

3 Application of the CLT, martingales, invariance principle

For a large class of dynamical systems of hyperbolic type, the method introduced by M. Gordin [6] gives a way to reduce, up to a regular coboundary, a Hölderian function Φ to a function satisfying a martingale condition. This allows to prove for regular functions which are not coboundaries, not only a CLT, but also a CLT for subsequences of positive density and the functional CLT (or the invariance principle).

In this subsection, we recall some results for martingale increments and briefly mention their application to find the set of essential directions..

3.1 Martingale methods and essential limit directions

The theorem of Ibragimov and Billingsley (see [2]) stated in terms of dynamical systems gives a CLT which can be extended to several improvements:

Proposition 3.1. *Let $(X, \mathcal{A}, \mu, T)$ be an ergodic invertible dynamical system and $\mathcal{F}$ a sub σ-algebra of $\mathcal{A}$ such that $\mathcal{F} \subset T^{-1}\mathcal{F}$. Let Φ be a $\mathbb{R}^d$-valued square integrable function, $\mathcal{F}$-measurable and such that the sequence $(\Phi \circ T^n)_{n \in \mathbb{Z}}$ is a sequence of martingale increments with respect to $(T^{-n}\mathcal{F})$ (equivalently by stationarity: $\mathbb{E}(\Phi | T\mathcal{F}) = 0$).*

If Φ is non contained a.s. in a fixed hyperplane, the cocycle (Φ_n) is such that $(\frac{1}{\sqrt{n}}\Phi_n)_{n\geq 1}$ has asymptotically a Gaussian law, with a non degenerated covariance matrix Γ.

For every strictly increasing sequence of measurable functions $(k_n)_{n\geq 1}$ with values in $\mathbb{N}$ such that, for a constant $a \in]0, \infty[$, $\lim_n k_n(x)/n = a$ exists a.e. we have:

$$\frac{1}{\sqrt{n}}\Phi_{k_n(.)}(.) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, a^{-1}\Gamma\right).$$

Moreover the cocycle (Φ_n) satisfies the invariance principle.

Theorem 3.2. *$\mathcal{ED}(\Phi) = \mathbb{S}^{d-1}$ under the conditions of the previous proposition.*

Proof. Let B be a subset of positive measure and let $(R_n(x))$ be the sequence of visit times in B. The induced cocycle (Φ_n^B) is obtained by sampling the cocycle (Φ_n) at the random times R_n of visits to B.
We have (Kac lemma): $\lim_n R_n(x)/n = 1/\mu(B)$, since by the ergodic theorem:

$$\frac{n}{R_n(x)} = \frac{1}{R_n(x)} \sum_{j=0}^{R_n(x)-1} 1_B\left(T^j x\right) \to \mu(B).$$

Therefore (Φ_n^B) satisfies the CLT, with the same covariance matrix as for the cocycle (Φ_n) up to a scalar. By Lemma 2.14, this implies the result. □

If the covariance matrix is degenerated, the set $\mathcal{ED}(\Phi)$ is the unit sphere of a subgroup isomorphic to $\mathbb{R}^{d'}$, for $d' < d$.

Reduction by cohomology to martingale increments

When Gordin's method can be applied, using Theorem 3.2 and the fact that the set of essential limit directions is the same for two cohomologous cocycles, we obtain $\mathcal{ED}(\Phi) = \mathbb{S}^{d'-1}$, for some $d' \leq d$, if Φ has values in $\mathbb{R}^d$.

This method can be used for Hölderian functions in many systems, among which: piecewise continuous expansive maps of the interval, toral automorphisms, geodesic and diagonal flows on homogeneous spaces of finite volume, dispersive billiards in the plane.

Let us give an explicit example.

Proposition 3.3. *Let T be an ergodic endomorphism of the torus $\mathbb{T}^r$, $r \geq 1$, endowed with the Lebesgue measure. If Φ is a Hölderian function with values in $\mathbb{R}^d$, then $\mathcal{ED}(\phi) = \mathbb{S}^{d'-1}$, for $d' \leq d$.*

Proof. The function Φ is cohomologous to Ψ such that $(\Psi \circ T^n)$ is a sequence of vector d'-dimensional martingale increments (with $d' \leq d$) (cf. [10]). We can apply Proposition 3.1, then Theorem 3.2. □

The situation for the models where Gordin's method is available is comparable to that of cocycles which are regular in the sense of Definition 2.2.

In the last section we will deduce a stronger property from the invariance principle.

3.2 Invariance principle and behavior of the directional process

Let (X, μ, T) be an ergodic dynamical system and let Φ be a measurable function on X with values in $\mathbb{R}^d$, $d \geq 2$. In this subsection, Φ is assumed to be bounded and centered.

We are going to give conditions on (Φ_n) which imply the property (2.4) introduced in the first section for the directional process $Z_n = \tilde{\Phi}_n(.) = \Phi_n(.)/\|\Phi_n(.)\|$.

We denote by $(W_n^\Phi)_{n\geq 1}$ (or simply $(W_n)_{n\geq 1}$) the interpolated piecewise affine process with continuous paths defined for $x \in X$ and $n \geq 1$ by

$$W_n^\Phi(x, s) = \Phi_k(x) + (ns - k)\Phi\left(T^k x\right) \text{ if } s \in \left[\frac{k}{n}, \frac{k+1}{n}\right].$$

If C is a cone with non empty interior $\overset{\circ}{C}$ and boundary ∂C of measure 0, the amount of time spent by $W_n^\Phi(x, s)$ in C is

$$\tau_{n,C}^\Phi(x) = \int_0^1 \mathbf{1}_C\left(W_n^\Phi(x, s)\right) ds.$$

Recall that the *invariance principle* for Φ, sometimes called Donsker's invariance principle, means here that the process $(n^{-\frac{1}{2}} W_n^\Phi(x, .))_{n\geq 1}$ (defined on the probability space (X, μ) and with values in the space $(\mathcal{C}_d([0, 1], \|\ \|_\infty)$ of continuous functions from $[0, 1]$ to $\mathbb{R}^d$ endowed with the uniform norm) converges in distribution to the standard Brownian motion in $\mathbb{R}^d$ (cf. [2]).

As mentioned before, the invariance principle, a by-product of the martingale method, is valid for large classes of regular functions in many dynamical systems of hyperbolic type.

Theorem 3.4. *Suppose that (X, μ, T) is ergodic, that the invariance principle is satisfied for a centered bounded function $\Phi : X \to \mathbb{R}^d$ and that C is a cone with non empty interior in $\mathbb{R}^d$ and with a complement with non empty interior. Then, for almost every x,*

$$\limsup_{n\to\infty} \tau_{n,C}^\Phi(x) = 1 \text{ and } \liminf_{n\to\infty} \tau_{n,C}^\Phi(x) = 0$$

We need preliminary lemmas before the proof of Theorem 3.4. Firstly, let us remark that the property stated in the theorem holds for the Brownian motion.

Visit of the Brownian motion in cones

Let $B = (B_s)$ denote the standard Brownian motion in $\mathbb{R}^d$, for $d \geq 2$. Consider a cone C with non empty interior $\overset{\circ}{C}$ in $\mathbb{R}^d$ and with a complement with non empty interior. The amount of time spent by (B_s) in C during the interval $[0, t]$ is $\tau_C(t) = \int\limits_0^t \mathbf{1}_C(B_s)\, ds$.

Proposition 3.5. *We have a.s.* $\limsup_{t\to\infty} \frac{1}{t}\tau_C(t) = 1$ *and* $\liminf_{t\to\infty} \frac{1}{t}\tau_C(t) = 0$.

Proof. Since the variable $\limsup_{t\to\infty} \frac{1}{t}\tau_C(t)$ is asymptotic, it is a.s. equal to a constant value $\ell \in [0, 1]$. Because of the scaling property of the Brownian motion and because C is a cone, we have

$$\begin{aligned}
\mathbb{P}\left(\frac{1}{t}\tau_C(t) \in I\right) &= \mathbb{P}\left(\left[\int\limits_0^t \mathbf{1}_C\left(B_s\right) \frac{ds}{t}\right] \in I\right) = \mathbb{P}\left(\left[\int\limits_0^1 \mathbf{1}_C\left(B_{ts}\right)\, ds\right] \in I\right) \\
&= \mathbb{P}\left(\left[\int\limits_0^1 \mathbf{1}_C\left(\sqrt{t}B_s\right)\, ds\right] \in I\right) = \mathbb{P}\left(\left[\int\limits_0^1 \mathbf{1}_C\left(B_s\right)\, ds\right] \in I\right).
\end{aligned}$$

Take $\alpha \in (0, 1)$. As the cone C has a non empty interior, we have $\mathbb{P}(B_\alpha \in \overset{\circ}{C}) > 0$ and, knowing that B_α is in $\overset{\circ}{C}$, we also have $\mathbb{P}(B_s \in \overset{\circ}{C},\ \forall s \in (\alpha, 1)) > 0$. The obvious inequality

$$\mathbb{P}\left(\frac{1}{t}\tau_C(t) > 1-\alpha\right) \geq \mathbb{P}\left(B_s \in \overset{\circ}{C},\ \forall s \in (\alpha, 1)\right)$$

then implies

$$\mathbb{P}\left(\frac{1}{t}\tau_C(t) > 1-\alpha\right) > 0,\ \forall \alpha > 0. \tag{3.1}$$

We have $\limsup_t \mathbb{P}(\frac{1}{t}\tau_C(t) > \ell+\varepsilon) \leq \mathbb{P}(\limsup_t(\frac{1}{t}\tau_C(t)) > \ell+\varepsilon) = 0, \forall \varepsilon > 0$.

Now the distribution of $\frac{1}{t}\tau_C(t)$ does not depend on t, so that we have $\mathbb{P}(\frac{1}{t}\tau_C(t) > \ell+\varepsilon) = 0$. In view of (3.1), this implies that $\ell+\varepsilon > 1-\alpha$. But α and ε being arbitrary small, one gets $\ell \geq 1$, that is $\ell = 1$. By considering the complement, we obtain the result for lim inf. □

This suggests that, if we can approximate our process (W_n^Φ) by a Brownian motion, then the property claimed in Theorem 3.4 holds. This is the case, for example if we can assert that for every $\gamma > 1/4$, there exists $C > 0$, so that, for all $t \in [0, 1]$, one has a.s.

$$\left\|B(nt) - W_n^\Phi(t)\right\| \leq Cn^\gamma.$$

Such a property is sometimes called an almost sure invariance principle. It has been established for some hyperbolic or quasi-hyperbolic systems (see Gouëzel [7]). To deduce the desired property for W_n from the one satisfied by the Brownian motion, we need to control the amount of time spent by the Brownian motion not too far from the origin and to enlarge or shrink the cone we are interested in to get convenient estimates. We will not do these computations here because they are very similar to what is done below.

Indeed, we will show that the "plain" Donsker's invariance principle suffices. From the preceding proof for the Brownian motion, we just keep in mind that (3.1) in Proposition 3.5 is true.

We need to know that, most of the time, W_n is far from the origin:

Lemma 3.6. *If Φ is not a coboundary, for every $M > 0$, the asymptotic frequency of visits of the process $(W_n)_{n\geq 1}$ to the ball $B(0, M)$ with center at the origin and radius $M > 0$ in $\mathbb{R}^d$ is almost surely zero:*

$$\lim_n \int_0^1 \mathbf{1}_{B(0,M)} \left(W_n(x,s)\right) \, ds = 0, \text{ for a.e. } x. \tag{3.2}$$

Proof. For $K > 0$, the ergodic theorem applied to $(X \times \mathbb{R}^d, T_\Phi, \lambda = \mu \times dy)$ and $\mathbf{1}_{B(0,K)}$ ensures the existence for a.e. $(x, y) \in X \times \mathbb{R}^d$ of the limit

$$u_K(x,y) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{B(0,K)} \left(\Phi_k(x) + y\right).$$

The function u_K is integrable on $X \times \mathbb{R}^d$, non negative and T_Φ-invariant. Suppose that $u_K \neq 0$ on a set of positive measure. Then $u_K\lambda$ is a finite T_Φ-invariant measure on $X \times \mathbb{R}^d$, absolutely continuous with respect to λ. Since (X, T, μ) is ergodic, this implies that Φ is a coboundary [3], contrary to the assumption. Therefore $u_K = 0$ a.e. for the measure λ.
Taking $K = M+1$, since $\mathbf{1}_{B(0,M+1)}(\Phi_k(x)+y) \geq \mathbf{1}_{B(0,M)}(\Phi_k(x))$, for $\|y\| \leq 1$, this implies:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} 1_{B(0,M)} \left(\Phi_k(x)\right) = 0, \text{ for a.e. } x.$$

Now we compare the discrete sum with the integral:

$$\int_0^1 \mathbf{1}_{B(0,M)} \left(W_n(x,s)\right) \, ds = \frac{1}{n} \sum_{k=0}^{n-1} \int_k^{k+1} \mathbf{1}_{B(0,M)} \left(\Phi_k(x) + (t-k)\Phi\left(T^k x\right)\right) \, dt.$$

Let $\varepsilon > 0$ be arbitrary and K be such that $\mu(|\Phi| > K) \leq \varepsilon$. We have for $t \in [k, k+1]$:

$$\mathbf{1}_{B(0,M)} \left(\Phi_k(x) + (t-k)\Phi\left(T^k x\right)\right) \leq \mathbf{1}_{|\Phi(T^k x)|>K} + \mathbf{1}_{B(0,M+K)} \left(\Phi_k(x)\right),$$

so that for a.e. x.

$$\begin{aligned} &\limsup_n \int_0^1 \mathbf{1}_{B(0,M)} \left(W_n(x,s)\right) \, ds \\ &\leq \limsup_n \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{B(0,M+K)} \left(\Phi_k(x)\right) + \lim_n \frac{1}{n} \sum_{k=0}^{n-1} \mathbf{1}_{(|\Phi|>K)} \left(T^k x\right) \leq 0 + \varepsilon. \end{aligned}$$

It implies (3.2). $\square$

Notation 3.7. Let us take $a > 0$ and $\underline{u}$ in the unit sphere in $\mathbb{R}^d$. For a and $\underline{u}$ fixed, for every $t > 0$ we denote by $C_{a,t}$ or simply C_t the cone $\{v \in \mathbb{R}^d : \|\frac{v}{\|v\|} - \underline{u}\| < at\}$.

Lemma 3.8. *Suppose Φ is not a coboundary. The set of discontinuity points of the increasing function $t \mapsto \limsup_{n\to\infty} \tau_{n,C_t}(x)$ is a.e. constant (with respect to x). If t does not belong to this set of discontinuity points, then $\limsup_{n\to\infty} \tau_{n,C_t}(x)$ is almost surely constant in x.*

Proof. Let us compare $\tau_{n,C_t}(x)$ and $\tau_{n,C_t}(Tx)$. Take $\varepsilon > 0$. For every $k \geq 1$, we have $\Phi_k(Tx) = \Phi_k(x) - \Phi(x) + \Phi(T^k x)$ and $\|\Phi_k(Tx) - \Phi_k(x)\| \leq 2\|\Phi\|_\infty$. There is M such that, if $\Phi_k(x) > M$ and $\Phi_k(x) \in C_t$, then $\Phi_k(Tx) \in C_{t+\varepsilon}$.
Therefore, for $s \in [0,1]$, we have $W_n(Tx,s) \in C_{t+\varepsilon}$, when $W_n(x,s) \in C_t$ and $W_n(Tx,s) \geq M$. This implies:

$$\begin{aligned}\int_0^1 \mathbf{1}_{C_t}\left(W_n(x,s)\right)\,ds &= \int_0^1 \mathbf{1}_{C_t\cap B(0,M)}\left(W_n(x,s)\right)\,ds + \int_0^1 \mathbf{1}_{C_t\cap B(0,M)^c}\left(W_n(x,s)\right)\,ds\\ &\leq \int_0^1 \mathbf{1}_{B(0,M)}\left(W_n(x,s)\right)\,ds + \int_0^1 \mathbf{1}_{C_{t+\varepsilon}}\left(W_n(Tx,s)\right)\,ds.\end{aligned}$$

When n tends to infinity, the first integral tends to 0 almost surely by (3.2) (Lemma 3.6) if Φ is not a coboundary. It follows:

$$\limsup_n \tau_{n,C_t}(x) \leq \limsup_n \tau_{n,C_{t+\varepsilon}}(Tx).$$

In the same way, we have $\limsup_n \tau_{n,C_t}(Tx) \leq \limsup_n \tau_{n,C_{t+\varepsilon}}(x)$. It follows, for every positive real numbers $s < t < u < v$:

$$\limsup_{n\to\infty} \tau_{n,C_s}(x) \leq \limsup_{n\to\infty} \tau_{n,C_t}(Tx) \leq \limsup_{n\to\infty} \tau_{n,C_u}(x) \leq \limsup_{n\to\infty} \tau_{n,C_v}(Tx).$$

This implies $\lim_{s\to t, s>t} \limsup_{n\to\infty} \tau_{n,C_s}(x) = \lim_{s\to t,s>t} \limsup_{n\to\infty} \tau_{n,C_s}(Tx)$, for every t.

Thus the common limit defines a map from X into the set of increasing functions on $[0,1]$. This map is T-invariant, hence almost surely constant by ergodicity of T. In particular the finite or countable set of discontinuity points of $t \mapsto \limsup_{n\to\infty} \tau_{n,C_t}(x)$ is independent of x. Outside this at most countable set of values of t, $\limsup_{n\to\infty} \tau_{n,C_t}(x)$ does not depend on x. ☐

Let $A_0(C) := \{\chi \in \mathcal{C}_d([0,1]) : \int_0^1 \mathbf{1}_{\partial C}(\chi(s))\,ds > 0\}$ be the set of functions taking their values in the boundary of C for a set of positive measure of the variable s.

Lemma 3.9. *If the Lebesgue measure of the boundary of C is zero, then the Wiener measure of the set $A_0(C)$ is 0.*

Proof. The Wiener measure of $A_0(C)$ is

$$W\left(A_0(C)\right) = \mathbb{P}((B_.) \in A_0(C)) = \mathbb{P}\left(\left\{\omega : \int_0^1 \mathbf{1}_{\partial C}\left(B_s(\omega)\right)\, ds > 0\right\}\right).$$

From the assumption on C, we have $\mathbb{P}(B_s \in \partial C) = 0$ for every s and therefore $\mathbb{E}\left(\int_0^1 \mathbf{1}_{\partial C}(B_s)\, ds\right) = \int_0^1 \mathbb{P}(B_s \in \partial C)\, ds = 0.$ □

Remark 1. It is clear that the set Δ of atoms of the distribution of $\int_0^1 \mathbf{1}_C(B_s)\, ds$ (the image probability on $[0, 1]$ of the Wiener measure on $\mathcal{C}_d([0, 1])$) is at most countable. If $c \notin \Delta$, then the set of functions χ in $\mathcal{C}_d([0, 1])$ such that $\int_0^1 \mathbf{1}_C(\chi(s))\, ds = c$ has zero measure for the Wiener measure.

For $\eta > 0$, denote by $\partial C(\eta)$ the set of points in $\mathbb{R}^d$ at a distance $\leq \eta$ from the boundary of C.

Proof of Theorem 3.4: Since the interior of C is non empty, there is a family of cones $C_t = C_{a,t}$ contained in C, constructed like the cones introduced in the preceding lemmas. We take a real $t > 0$ such that $\limsup_{n\to\infty} \tau_{n,C_t(x)}$ is almost surely constant in x. If we show that $\limsup_{n\to\infty} \tau_{n,C_t} = 1$, then also $\limsup_{n\to\infty} \tau_{n,C} = 1$.

From now, on we replace C by C_t still denoted C. In particular, the boundary of C now has Lebesgue measure 0. The invariance principle "$W_n(.)/\sqrt{n} \to B_.$" means that, for every continuous functional F on $\mathcal{C}_d([0, 1])$, we have

$$\mathbb{E}\left(F\left(\frac{W_n(*, .)}{\sqrt{n}}\right)\right) \to \mathbb{E}\left(F(B_.)\right). \tag{3.3}$$

Suppose that a sequence of probability measures $(\mathbb{P}_n)$ defined on a space X converges weakly to $\mathbb{P}$. By Theorem 2.7 in Billingsley's book [2], if a measurable function Ψ from X to a metric space Y has a set of discontinuity points of measure zero for $\mathbb{P}$, then the sequence of pushforward measures $(\mathbb{P}_{n,\Psi})$ converges to the pushforward measure $\mathbb{P}_\psi$.

For Ψ we take here the function F_c, $c > 0$, defined on the metric space of continuous functions from $[0, 1]$ to $\mathbb{R}^d$ with the uniform norm by

$$F_c(\chi) = \mathbf{1}_{[c,+\infty]}\left(\int_0^1 \mathbf{1}_C\left(\chi(s)\right)\, ds\right). \tag{3.4}$$

In order to apply to F_c the quoted theorem and the convergence (3.3), we have to show that the set of discontinuity points of F_c has measure zero for the Wiener measure.

Assume that $c \notin \Delta$ (i.e., c is not an atom of the distribution of $\int_0^1 \mathbf{1}_C(B_s)\, ds$). Let us consider the set $\mathcal{G}_c$ of functions χ_0 such that $\chi_0 \notin A_0(C)$, i.e., the set $\{t : \chi_0(t) \in \partial C\}$ has Lebesgue measure 0, and $\int_0^1 \mathbf{1}_C(\chi_0(s))\, ds \neq c$. It has full Wiener measure by Lemma 3.9 and Remark 1.

Let us show that F_c is continuous on the set $\mathcal{G}_c$.

Let ε be such that $0 < \varepsilon < |\int_0^1 \mathbf{1}_C(\chi_0(s))\,ds - c|$. The measure of the set of times s for which $\chi_0(s)$ is at a distance less than η from the boundary of C tends to 0 when η tends to 0. We can take $\eta > 0$ such that this measure is less then ε.

Let χ be another function at a uniform distance less than η from χ_0. If $\chi_0(s)$ is not in $\partial C(\eta)$, then $\chi_0(s)$ and $\chi(s)$ are either both in C^c or both in C. Thus, we have

$$\begin{aligned}
&\left|\int_0^1 \mathbf{1}_C\left(\chi_0(s)\right)\, ds - \int_0^1 \mathbf{1}_C\left(\chi(s)\right)\, ds\right| \\
&\le \left[\int_0^1 \mathbf{1}_{\partial C(\eta)}\left(\chi_0(s)\right) + \mathbf{1}_{(\partial C(\eta))^c}\left(\chi_0(s)\right)\right] \left|\mathbf{1}_C\left(\chi_0(s)\right) - \mathbf{1}_C\left(\chi(s)\right)\right| ds \\
&\le \int_0^1 \mathbf{1}_{\partial C(\eta)}\left(\chi_0(s)\right)\, ds + \int_0^1 \mathbf{1}_{(\partial C(\eta))^c}\left(\chi_0(s)\right) \left|\mathbf{1}_C\left(\chi_0(s)\right) - \mathbf{1}_C\left(\chi(s)\right)\right| ds \le \varepsilon + 0.
\end{aligned}$$

Therefore: $\mathbf{1}_{[c,+\infty[}\Big(\int_0^1 \mathbf{1}_C(\chi(s))\,ds\Big) = \mathbf{1}_{[c,+\infty[}\Big(\int_0^1 \mathbf{1}_C(\chi_0(s))\,ds\Big)$ and we have proved that the functional F_c is continuous at χ_0.

Finally we have shown, for every $c \notin \Delta$, the continuity of F_c on the set $\mathcal{G}_c$ which has full Wiener measure.

For c outside Δ (which is at most countable), it follows from the theorem mentioned above:

$$\mathbb{E}\left(\frac{F_c\left(W_n(*,.)\right)}{\sqrt{n}}\right) \to \mathbb{E}\left(F_c(B_.)\right),$$

that is:

$$\lim_n \mathbb{P}(x : \left[\int_0^1 \mathbf{1}_C\left(\frac{W_n(x,s)}{\sqrt{n}}\right)\, ds\right] \ge c) = \mathbb{P}\left(\left[\int_0^1 \mathbf{1}_C\left(B_s\right)\, ds\right] \ge c\right). \tag{3.5}$$

As C is a cone, we have:

$$\int_0^1 \mathbf{1}_C\left(\frac{W_n(x,s)}{\sqrt{n}}\right)\, ds = \tau_{n,C}(x).$$

Because of (3.1), (3.5) implies that, for every $c < 1$ with $c \notin \Delta$, $\mathbb{P}(\tau_{n,C} \ge c) > 0$ for n large enough. As a consequence, we have

$$\mathbb{P}(\limsup_n \tau_{n,C} \ge c) \ge \limsup_n \mathbb{P}(\tau_{n,C} \ge c) > 0.$$

Hence, $\limsup_n \tau_{n,C}$ being constant, it follows $\limsup_n \tau_{n,C} \ge 1$. As $\tau_{n,C} \in [0,1]$, this proves $\limsup_n \tau_{n,C} = 1$. □

Remark 2. We can also consider the piecewise constant function: $V_n(x,s) := \Phi_k(x)$ for $s \in [k/n, (k+1)/n[$. If Φ is bounded, then

$$\left\| W_n(x,\cdot) - V_n(x,\cdot)\right\|_\infty \leq \|\Phi\|_\infty .$$

On the other hand, we have

$$\int_0^1 \mathbf{1}_C\left(V_n(x,s)\right)\, ds = \frac{1}{n}\int_0^n \mathbf{1}_C\left(V_n\left(x, \frac{t}{n}\right)\right) dt = \frac{1}{n}\mathrm{Card}\left\{k \leq n : \ \Phi_k(x) \in C\right\}.$$

Reasoning as before we can show that if a cone C contains a cone of the form C_t, like in Lemma 3.8, then

$$\limsup_n \int_0^1 \mathbf{1}_C\left(V_n(x,s)\right)\, ds \geq \limsup_n \int_0^1 \mathbf{1}_{C_t}\left(W_n(x,s)\right)\, ds = 1.$$

This means that, if Φ is a bounded function satisfying Donsker's invariance principle, we also have the following discrete version of the property claimed in the theorem:

$$\limsup_n \frac{1}{n}\mathrm{Card}\left\{k \leq n : \ \Phi_k(x) \in C\right\} = 1.$$

Acknowledgement The authors thank B. Delyon and S. Gouëzel for fruitful discussions. The authors are grateful to the referee for his careful reading.

References

[1] Aaronson (J.), *An introduction to infinite ergodic theory*, Mathematical surveys and monographs 50, AMS, Providence, 1997.

[2] Billingsley (P.), *Convergence of probability measures*. Second edition. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley and Sons, Inc., New York, 1999.

[3] Conze (J.-P.), Transformations cylindriques et mesures finies invariantes, Ann. Sci. Univ. Clermont Math. **17** (1979), 25–31.

[4] Conze (J.-P.), Gutkin (E.), On recurrence and ergodicity for geodesic flows on non-compact periodic polygonal surfaces, Ergodic Theory Dynam. Systems, 32 (2012), no. 2, 491–515.

[5] Derriennic (Y.), Ergodic theorem, reversibility and the filling scheme, Colloq. Math. 118 (2010), no. 2, 599–608.

[6] Gordin (M. I.): The central limit theorem for stationary processes, (Russian) Dokl. Akad. Nauk SSSR 188 (1969), 739–741.

[7] Gouëzel (S.), Almost sure invariance principle for dynamical systems by spectral methods, Ann. Probab. 38 (2010), no. 4, 1639–1671.

[8] Gutkin (E.), Geometry, topology and dynamics of geodesic flows on non-compact polygonal surfaces, Regul. Chaotic Dyn. 15 (2010), 482–503.

[9] Kesten (H.), Sums of stationary sequences cannot grow slower than linearly, Proc. Amer. Math. Soc. 49 (1975), 205–211.

[10] Le Borgne (S.), Limit theorems for non-hyperbolic automorphisms of the torus, Israel J. Math. 109 (1999), 61–73.
[11] Pène (F.), Applications des propriétés stochastiques du billard dispersif. (French) [Consequences of the stochastic properties of dispersed billiards] C.R.A.S. Paris Sér. I Math. 330 (2000), no. 12, 1103-1106.
[12] Schmidt (K.), *Lectures on Cocycles of Ergodic Transformations Groups*, Lect. Notes in Math. vol. 1, Mac Millan Co. of India (1977).
[13] Szász (D.), Varjú (T.), Local limit theorem for the Lorentz process and its recurrence in the plane. Ergodic Theory Dynam. Systems 24 (2004), no. 1, 257–278.
[14] Tanny (D.), A zero-one law for stationary sequences, Z. Wahrsch. und Verw. Gebiete 30 (1974), 139–148.
[15] Woś, J., The filling scheme and the ergodic theorems of Kesten and Tanny. Colloq. Math. 52 (1987), no. 2, 263–276.

Joseph Rosenblatt

Optimal norm approximation in ergodic theory

Abstract: Given an ergodic transformation τ, and a mean-zero $f \in L_2(X)$, the ergodic averages $A_n^\tau f = \frac{1}{n}\sum_{k=1}^n f \circ \tau^k$ converge in L_2-norm to zero. However, for a fixed value of n, there could be other powers $m_1, \dots, m_n$ such that $\|\frac{1}{n}\sum_{k=1}^n f \circ \tau^{m_k}\|_2$ is much smaller than $\|\frac{1}{n}\sum_{k=1}^n f \circ \tau^k\|_2$. For specific functions and transformations, we seek to compute, or estimate, the infimum of the norms $\|\frac{1}{n}\sum_{k=1}^n f \circ \tau^{m_k}\|_2$. This allows us to show that for the generic dynamical systems, given in addition the generic function, the usual ergodic averages are infinitely often very far from giving the optimal L_2-norm approximation among averages of the form $\frac{1}{n}\sum_{k=1}^n f \circ \tau^{m_k}$, and yet at the same time, again generically, the usual ergodic averages are infinitely often very close to giving the optimal L_2-norm approximation among averages of the form $\frac{1}{n}\sum_{k=1}^n f \circ \tau^{m_k}$.

Keywords: Ergodic, Averages, Optimal.

1 Introduction

Take an invertible ergodic measure-preserving transformation τ of a standard probability space (X, β, p). The averages $A_n^\tau f = \frac{1}{n}\sum_{k=1}^n f \circ \tau^k$ converge in L_r-norm for any $f \in L_r(X), 1 \leq r < \infty$. But for a fixed value of n and a fixed mean-zero f, we could potentially get closer to 0, which is the L_r-norm limit of $A_n^\tau f$, if we used instead an average of the form $\frac{1}{n}\sum_{k=1}^n f \circ \tau^{m_k}$ for some $m_1, \dots, m_n$. So for $f \in L_r(X)$ of mean-zero, let $O_{n,r}^\tau f = \inf\{\|\frac{1}{n}\sum_{k=1}^n f \circ \tau^{m_k}\|_r : m_1, \dots, m_n \in \mathbb{Z}\}$. We would like to determine for particular transformations and functions the size of this optimal lower bound. It may be worthwhile to look at the potentially smaller approximation $C_n^\tau f$ one gets by taking convex combinations of n powers. That is, $C_{n,r}^\tau f = \inf\{\|\sum_{k=1}^n \lambda_k f \circ \tau^{m_k}\|_r : \lambda_k \geq 0, \sum_{k=1}^n \lambda_k = 1, \text{ and } m_1, \dots, m_n \in \mathbb{Z}\}$. For now we focus on $O_{n,r}^\tau f$. See the remark at the end of the article for a short discussion of some of the issues here.

The Mean Ergodic Theorem itself implies that $O_{n,r}^\tau f$ tends to zero as n tends to ∞ for all mean-zero $f \in L_r(X)$. But one might expect that in fact the rate that $O_{n,r}^\tau f$ tends to zero can far exceed the rate at which $\|A_n^\tau f\|_r$ tends to zero. Here are some examples of questions that we focus on in this text.

1. For which transformations and functions do we have $O_{n,r}^\tau f = 0$ for some, or all sufficiently large n?

2. Is there a quantitative rate that $O^\tau_{n,r}f$ goes to zero? How does it depend on the transformation and the function?
3. In terms of f, τ, and n, when is $O^\tau_{n,r}f$ significantly smaller than the norm $\|A^\tau_n f\|_r$?
4. In contrast, in terms of f, τ, and n, when is the norm $\|A^\tau_n f\|_r$ close to being equal to $O^\tau_{n,r}f$?

It will be convenient to use the spectral measures ν^τ_f of the dynamical system. These are defined as usual for any $f \in L_2(X)$ by $\widehat{\nu^\tau_f}(n) = \langle f, f \circ \tau^n\rangle = \int_X f\overline{f \circ \tau^n}\, dp$ for all n. The measures ν^τ_f are positive Borel measures with $\|\nu^\tau_f\|_1 = \|f\|_2^2$.

In addition, we will want to use discrete measures as operators on $L_r(X)$. So if μ is a finitely supported discrete measure $\sum_k d_k\delta_k$, then we define $\mu^\tau f = \sum_k d_k f \circ \tau^k$ for all $f \in L_r(X)$. Also, we take the Fourier transform of μ to be given by $\widehat{\mu}(\gamma) = \sum_k d_k\gamma^k$ for all $\gamma \in \mathbb{T}$. We write μ_n when μ has n-terms. We reserve the notation $A^\tau_{n,r}$ for the usual ergodic averages, which is the particular case of μ^τ_n where $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_k$.

There are a number of interesting issues that arise in studying optimal averages. We will first consider these in $L_2(X)$ where we can obtain the best results. Then at the end we will look at other Lebesgue spaces. For simplicity, we write $O^\tau_n f$ to denote $O^\tau_{n,2}f$ for $f \in L_2(X)$. We prove a number of Baire category results here. By a *residual* set, we will mean the complement of a set of first category. Such a set necessarily contains a G_δ set. In all of the cases in this paper, the residual sets are dense, and hence not empty. Also, we say that a property holds generically when there is a residual set on which it holds.

We should also observe that we focus on structure results for the Koopman operators $Tf = f \circ \tau$. However, many of the results hold for a wider class of operators, either Markov operators or, even more generally, contraction operators. We will not try to state the results in full generality, but when it seems particularly worthwhile we will point out when a result could be proved without much change in the argument for a wider class of operators. In the same vein, many of the results can be stated and proved for actions of groups by measure-preserving transformations. This generality will be left for some future work.

2 Basics

It is clear that we can sometimes have $O^\tau_n f = 0$. For example, take f such that $f \circ \tau = -f$, i.e., f is an eigenvector with eigenvalue -1 for the Koopman operator given by τ. Then $A^\tau_2 f = 0$ and so of course $O^\tau_2 f$. This use of eigenvalues is discussed more in later parts of this paper. For now, suffice it to say that for discrete spectrum maps we can have $O^\tau_n f = 0$ when f is an eigenvector for τ, and also without f being an eigenvector for τ. We can also have $O^\tau_n f = 0$ for particular f even though τ is weakly mixing and there are no non trivial eigenvectors. However, as we prove below, for strongly mixing τ, we always have $O_n^\tau f > 0$ if f is not zero.

But, in general, we cannot have $O_n^\tau f = 0$ for all mean-zero functions $f \in L_2(X)$.

Proposition 2.1. *Let $f = 1_E - p(E)$. Then for $np(E) < 1$, and any $m_1, \dots, m_n$, we have*

$$\left\| \frac{1}{n} \sum_{k=1}^{n} f \circ \tau^{m_k} \right\|_2 \geq \frac{\sqrt{p(E)(1 - np(E))}}{\sqrt{n}}.$$

Proof. It is clear that $\| \sum_{k=1}^n f \circ \tau^{m_k} \|_2^2 = \int_X | \sum_{k=1}^n 1_E \circ \tau^{m_k} |^2 \, dp - n^2 p(E)^2$. By expanding the squared integrand and ignoring the correlations, we have $\int_X | \sum_{k=1}^n 1_E \circ \tau^{m_k} |^2 \, dp \geq np(E)$. Hence, $\| \sum_{k=1}^n f \circ \tau^{m_k} \|_2^2 \geq np(E) - n^2 p(E)^2 > 0$. This gives the underestimate. □

Corollary 2.2. *If $np(E) < 1$, then $O_n^\tau(1_E - p(E)) \geq \frac{\sqrt{p(E)(1-np(E))}}{\sqrt{n}}$.*

Remark 2.3. This shows the difference between considering a Koopman operator and a general unitary operator. For example, let H be a Hilbert space with an orthogonal basis $(e_n : n = 1, 2, 3, \dots)$. Let $U(e_n) = -e_n$ for all n. Then for all $f \in H$, we have $f + Uf = 0$. But we cannot have $f + Uf = 0$ for all mean-zero f in $L_2(X)$ when U is a Koopman operator given by a transformation τ.

More specifically,

Corollary 2.4. *If $np(E) < 1$, then $O_n^\tau(1_E - p(E)) \geq a_n \frac{\|1_E - p(E)\|_2}{\sqrt{n}}$, where $a_n = \sqrt{\frac{1-np(E)}{1-p(E)}}$.*

Remark 2.5. This result tells us something about homogeneous estimates of the form $O_n^\tau f \leq C_n \|f\|_2$. If this estimate is to hold for all mean-zero f, then by letting $p(E)$ tend to zero, we see that $C_n \geq \frac{1}{\sqrt{n}}$. This is an important point because we show in Proposition 4.2 that for τ weakly mixing, we always have $O_n^\tau f \leq \|f\|_2/\sqrt{n}$. So by Corollary 2.4, this overestimate cannot be globally improved.

Remark 2.6. It is clear that the above is the best underestimate we can give for functions f of the form $1_E - p(E)$. Indeed, consider the case of an ergodic τ, which is not totally ergodic. Suppose also that for some E, with $p(E) = 1/n$, we have $\tau^n E = E$ and $\tau^i E$, $i = 1, \dots, n$ pairwise disjoint. Then $\sum_{i=1}^n (1_E - 1/n) \circ \tau^i = 0$. So $O_n^\tau(1_E - 1/n) = 0$ too. This is reflective of computations in the next section because here τ has eigenfunctions with n-th roots of unity as eigenvalues. Indeed, let $\gamma = \exp(2\pi i/n)$ and $j = 1, \dots, n$. Let $\chi_j = \sum_{i=1}^n \gamma^{ij}(1_E - 1/n) \circ \tau^i$. Then $\chi_j \circ \tau = \gamma^{-j}\chi_j$ for all $j = 1, \dots, n$.

We can use our basic underestimate in Corollary 2.2 to show that $O_n^\tau f = 0$ in some sense rarely.

Proposition 2.7. *Let $\mathcal{Z}$ consist of all mean-zero functions $f \in L_2(X)$ such that $O_n^\tau f = 0$ for some $n \geq 2$. Then $\mathcal{Z}$ is a set of first category.*

Proof. Consider $\mathcal{Z}_n = \{f \in L_2(X) : \int f dp = 0, O_n^\tau f = 0\}$. We claim that $\mathcal{Z}_n$ is closed and has no interior. This proves the result since $\mathcal{Z} = \bigcup_{n=2}^\infty \mathcal{Z}_n$. First, suppose (f_s) is a sequence in $\mathcal{Z}_n$ and $\|f_s - f\|_2 \to 0$ as $s \to \infty$. Choose any s and $\epsilon > 0$. Then there is $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{m_k}$ such that $\|\mu_n^\tau f_s\|_2 \leq \epsilon$. We have

$$\begin{aligned}
\|\mu_n^\tau f\|_2 &\leq \|\mu_n^\tau(f - f_s)\|_2 + \|\mu_n^\tau f_s\|_2 \\
&\leq \|\mu_n^\tau(f - f_s)\|_2 + \epsilon \\
&\leq \|f - f_s\|_2 + \epsilon.
\end{aligned}$$

Hence, by first choosing s and then choosing μ_n, we can show that there exists μ_n such that $\|\mu_n^\tau f\|_2 \leq 2\epsilon$. Thus, $O_n^\tau f = 0$ too and $f \in \mathcal{Z}_n$.

Suppose $\mathcal{Z}_n$ contains nonempty interior. Then there exists $f_0 \in \mathcal{Z}_n$ and $\epsilon_0 > 0$ such that for all mean-zero g with $\|g\|_2 \leq 1$, we have $f_0 + \epsilon_0 g \in \mathcal{Z}_n$. Fix such a g and any $\delta > 0$. Then there are μ_n and ν_n such that $\|\mu_n^\tau f_0\|_2 \leq \delta$ and $\|\nu_n^\tau(f_0 + \epsilon_0 g)\|_2 \leq \delta$. But then

$$\begin{aligned}
\|(\mu_n * \nu_n)^\tau(\epsilon_0 g)\|_2 &\leq \|(\mu_n * \nu_n)^\tau(f_0 + \epsilon_0 g)\|_2 + \|(\mu_n * \nu_n)^\tau f_0\|_2 \\
&\leq \|\nu_n^\tau(f_0 + \epsilon_0 g)\|_2 + \|\mu_n^\tau f_0\|_2 \\
&\leq 2\delta.
\end{aligned}$$

Hence, $\|(\mu_n * \nu_n)^\tau g\|_2 \leq 2\delta/\epsilon_0$. Since δ is arbitrary, this shows that $O_{n^2}^\tau g = 0$. But we could have taken $g = (1_E - p(E))/\|1_E - p(E)\|_2$. For n fixed and $p(E)$ sufficiently small, we would have $O_{n^2}^\tau g > 0$. This contradiction shows that $\mathcal{Z}_n$ has no interior. □

3 Discrete Spectrum Case

Consider the case where τ is a discrete spectrum transformation. Assume that τ is ergodic, so the τ-invariant functions are constants and there is a complete orthonormal basis of eigenfunctions $(\chi_j : j \geq 0)$, where $\chi_j \circ \tau = \alpha_j \chi_j$ for a sequence of distinct $\alpha_j \in \mathbb{T}$. Assume $\chi_0 = 1$ and $\alpha_0 = 1$. It might seem that we could consider more generally a unitary operator U for which there is an orthogonal basis $(\chi_j \geq 0)$ of $L_2(X)$ which consists of eigenfunctions for U. The caveat is what was already observed in Remark 2.3.

Now take $f \in L_2(X)$ and write $f = \sum_j c_j \chi_j$. Then $f \circ \tau^m = \sum_j c_j \alpha_j^m \chi_j$. Using the discrete measure $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{m_k}$, we have $\mu_n^\tau f = \sum_j c_j \widehat{\mu}_n(\alpha_j)\chi_j$. We want to estimate this expression by controlling the value $\widehat{\mu}_n(\alpha_j)$ term by term.

First consider a prototype case: take just one $\alpha \neq 1$ in $\mathbb{T}$ and $f = \chi$ which is an eigenfunction for τ with eigenvalue α. Fix n. Assume that α is infinite order. Then we can choose m_0 so that α^{m_0} is ϵ close to a primitive n-th root of unity. Then let $m_k = km_0$, $k = 1, \ldots, n$. It follows that $\widehat{\mu}_n(\alpha)$ will be δ close to zero where $\delta \to 0$ as $\epsilon \to 0$. Thus, for this special case of f and α, we would have $O_n^\tau f = 0$.

We can use this idea to obtain the following extended result.

Proposition 3.1. *Suppose τ is an ergodic rotation of the circle given by $\tau(\gamma) = \alpha\gamma$ for all $\gamma \in \mathbb{T}$. Then for a dense subspace D of mean-zero functions $f \in L_2(\mathbb{T})$, there is $n(f) \geq 1$ such that for $n \geq n(f)$, we have $O_n^\tau f = 0$.*

Proof. Take a mean-zero trigonometric function $f_N = \sum_j c_j\chi_j$, where, $\chi_j(\gamma) = \gamma^j$ for all $\gamma \in \mathbb{T}$. The sum is taken over all j, $0 < |j| \leq N$. Then let $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{m_k}$. We have $\|\mu_n^\tau f_N\|_2^2 = \sum_j |c_j|^2 |\widehat{\mu}_n(\alpha^j)|^2$. So if $n > N$, and if we choose μ_n as before, we would have all of the terms α^{jm_0}, $0 < |j| \leq N$, also as close as we would like to non trivial n-th roots of unity. Hence, $\|\mu_n^\tau f_N\|_2^2$ can be made as small as we like with a suitable choice of m_0. This shows that $O_n^\tau f_N = 0$ when $n > N$. Thus, we can take our dense subspace D to consist of all the mean-zero trigonometric polynomials. □

Remark 3.2. Given τ in Proposition 3.1, and a fixed n, does there exist J_n such that for all $|j| \geq J_n$, we have $O_n^\tau \chi_j > 0$?

We can prove a similar result if τ is an ergodic rotation of a d-dimensional torus. Indeed, suppose $\tau(\gamma_1, \ldots, \gamma_d) = (\alpha_1\gamma_1, \ldots, \alpha_d\gamma_d)$ for all $(\gamma_1, \ldots, \gamma_d) \in \mathbb{T}^d$. Assume that τ is ergodic. This means that we are assuming $\alpha_i, i = 1, \ldots, d$ generates a free abelian group $\mathbb{Z}^d$ in $\mathbb{T}$. Also, given any particular $\theta_i, i = 1, \ldots, d$ and $\epsilon > 0$, there exists m_0 such that $|\alpha_i^{m_0} - \theta_i| \leq \epsilon$ for all $i = 1, \ldots, d$. It is then not hard to see that an argument similar to the one in Proposition 3.1 gives the same result for $\mathbb{T}^d$ in place of $\mathbb{T}$. Actually, this can even be carried out for the case of the infinite dimensional torus $\prod_{j=1}^\infty \mathbb{T}$ but now τ given by multiplication by $(\alpha_j : j \geq 1)$ with $(\alpha_j : 1 \leq j \leq J)$ generating an isomorphic copy of $\mathbb{Z}^J$ for all J.

Proposition 3.3. *Suppose τ is an ergodic rotation of $\prod_{j=1}^\infty \mathbb{T}$. Then for a dense subspace D of mean-zero functions $f \in L_2(\prod_{j=1}^\infty \mathbb{T})$, there is $n(f) \geq 1$ such that for $n \geq n(f)$, we have $O_n^\tau f = 0$.*

Remark 3.4. We cannot do quite as well as the above when α is a root of unity, say a p-th root of unity. However, we can get an estimate. Suppose n is a multiple of p, we could choose $m_0 = 1$ and so $m_k = k$ for $k = 1, \ldots, n$. It follows that we would have

$\sum_{k=1}^{n} \chi \circ \tau^{m_k} = 0$ exactly. For other values of n, we can write $n = mp + r$, $0 \le r < p$, and then see that $\sum_{k=1}^{n} f \circ \tau^{m_k} = \sum_{k=mp+1}^{n} \alpha^k \chi$. So with $f = \chi$, we have $O_n^\tau f \le \frac{p}{n}$.

Question I: Given τ with discrete spectrum, is it always the case that there is a dense subspace D of mean-zero functions $f \in L_2(\mathbb{T})$, there is $n(f) \ge 1$ such that for $n \ge n(f)$, we have $O_n^\tau f = 0$?

The answer to this is negative. For example, consider the case when τ has an eigenfunction with eigenvalue -1, i.e., $\chi \circ \tau = -\chi$. Then for $m_1, \dots, m_n$, and n odd, the smallest we can get for $\|\mu_n^\tau \chi\|_2$ is $\frac{1}{n}$. This means the dense subspace D cannot exist.

The one obvious fact we have is the following.

Proposition 3.5. *If τ has discrete spectrum and f is a mean-zero finite linear combination of characters, then for any N, there exists $n \ge N$ such that $O_n^\tau f = 0$.*

Proof. We can write $f = f_1 + f_2$ such that both f_i are also finite linear combinations of characters, but the associated eigenvalues for f_1 are roots of unity, and the associated eigenvalues for f_2 span a free abelian subgroup of $\mathbb{T}$. If n is large enough, then $\alpha^{(n-1)} = 1$ for all the roots of unity eigenvalues associated with f_1. Hence, $A_{(n-1)}^\tau f_1 = 0$. But for any $\epsilon > 0$, for sufficiently large n, we can also choose μ_n such that $\|\mu_n^\tau f_2\|_2 \le \epsilon$. Hence, $\|(A_{(n-1)} * \mu_n)^\tau (f_1 + f_2)\|_2 \le \epsilon$ too. Since $\epsilon > 0$ is arbitrary, this show $O_n^\tau f = 0$ for sufficiently large n. □

We can use the estimates above to show that in some sense there are not many mean-zero functions such that $\|A_n^\tau f\|_2 \le K O_n^\tau f$ for all sufficiently large values of n.

Proposition 3.6. *Suppose τ is an ergodic rotation of $\mathbb{T}^m$ where m is whole number, or the cardinal ω. Fix a subsequence $\mathbf{N}$ of $\mathbb{N}$ and a sequence (K_n). Let $\mathcal{O}(\mathbf{N})$ consist of all mean-zero functions such that for some K_n and N, we have $\|A_n^\tau f\|_2 \le K_n O_n^\tau f$ for all $n \ge N, n \in \mathbf{N}$. Then $\mathcal{O}(\mathbf{N})$ is a set of first category.*

Proof. We prove this for simplicity when m is a whole number. We are considering $\mathcal{O}(\mathbf{N}) = \bigcup_{N=1}^{\infty} \bigcap_{n \ge N, n \in \mathbf{N}} \mathcal{O}_n$ where $\mathcal{O}_n = \{f : \int f\, dp = 0, \|A_n^\tau f\|_2 \le K_n O_n^\tau f\}$. We claim that each $\mathcal{O}_n$ is closed, so $\bigcap_{n \ge N, n \in \mathbf{N}} \mathcal{O}_n$ is closed. But also, $\bigcap_{n \ge N, n \in \mathbf{N}} \mathcal{O}_n$ has no interior. Hence, $\mathcal{O}(\mathbf{N})$ is a set of first category.

Suppose (f_s) is in $\mathcal{O}_n$ and $\|f_s - f\|_2 \to 0$ as $s \to \infty$. Then for any μ_n, $\frac{1}{K_n}\|A_n^\tau f_s\|_2 \le \|\mu_n^\tau f_s\|_2$. So

$$\begin{aligned}
\frac{1}{K_n}\|A_n^\tau f\|_2 &\le \frac{1}{K_n}\|A_n^\tau (f - f_s)\|_2 + \frac{1}{K_n}\|A_n^\tau f_s\|_2 \\
&\le \frac{1}{K_n}\|f - f_s\|_2 + \|\mu_n^\tau f_s\|_2 \\
&\le \frac{1}{K_n}\|f - f_s\|_2 + \|f - f_s\|_2 + \|\mu_n^\tau f\|_2.
\end{aligned}$$

So letting $s \to \infty$, this shows that $\frac{1}{K_n}\|A_n^\tau f\|_2 \leq \|\mu_n^\tau f\|_2$ for all μ_n. Hence, $\|A_n^\tau f\|_2 \leq K_n O_n^\tau f$ and $f \in \mathcal{O}_n$.

Now consider the closed set $\bigcap_{n\geq N, n\in \mathbf{N}} \mathcal{O}_n$. If this set contains interior, then there is a non trivial trigonometric polynomial f, i.e., f is a finite linear combination of characters on $\mathbb{T}^m$, such that $f \in \bigcap_{n\geq N, n\in \mathbf{N}} \mathcal{O}_n$. But then there are values of $n \in \mathbf{N}$, sufficiently large such that $O_n^\tau f = 0$ and also $\|A_n^\tau f\|_2 \leq K_n O_n^\tau f$. So $A_n^\tau f = 0$ for some (large) value of n. This is not possible. Indeed, here τ is given by $\tau(\gamma_1,\dots,\gamma_d) = (\alpha_1\gamma_1,\dots,\alpha_d\gamma_d)$ for all $(\gamma_1,\dots,\gamma_d) \in \mathbb{T}^m$. The function $f = \sum_{\mathbf{l}} c_{\mathbf{l}}\chi_{\mathbf{l}}$, a finite linear combination of characters. Here the sum is over non zero indices $\mathbf{l} \in \mathbb{Z}^d$ and the coefficients $c_{\mathbf{l}}$ are assumed to be non zero. Also, given $\mathbf{l} = (l_1,\dots,l_d)$, the characters $\chi_{\mathbf{l}}(\gamma_1,\dots,\gamma_d) = \gamma_1^{l_1}\dots\gamma_d^{l_d}$. Hence, $n^2\|A_n^\tau f\|_2^2 = \sum_{\mathbf{l}} |c_{\mathbf{l}}|^2 |\sum_{k=1}^n (\alpha_1^{l_1}\dots\alpha_d^{l_d})^k|^2$. Let $\alpha = \alpha_1^{l_1}\dots\alpha_d^{l_d}$. Because $\mathbf{l}$ is not zero, and τ is ergodic, we have $\alpha \neq 1$. Also, $|\sum_{k=1}^n (\alpha_1^{l_1}\dots\alpha_d^{l_d})^k|^2 = |\frac{\alpha^{n+1}-\alpha}{\alpha-1}|^2$. So since also $\alpha^{n+1} \neq \alpha$, the multiplier $|\sum_{k=1}^n (\alpha_1^{l_1}\dots\alpha_d^{l_d})^k|^2$ is not zero for all $n \geq 2$. Thus, because f is non zero, we cannot have $A_n^\tau f = 0$. □

Remark 3.7. It is not clear if this can be improved so that the sequence $\mathbf{N}$ can depend on the function. That is, can we prove a category result for the mean-zero functions f such that there exists a constant K and a sequence $\mathbf{N}$ in $\mathbb{N}$ such that $\|A_n^\tau f\|_2 \leq K O_n^\tau f$ for $n \in \mathbf{N}$? The problem is that there are too many subsequences to allow a good Borel description of the functions in question. One could use the class of functions such that there exists K such that for all N, there is some $n \geq N$ such that $\|A_n^\tau f\|_2 \leq K O_n^\tau f$. However, it is not clear how to write this set as an F_σ set, or find an F_σ set that contains it, and then also show that the F_σ set is a set of first category.

Remark 3.8. It is not hard to see that G is metric compact abelian and monothetic if and only if its dual group is a countable subgroup of $\mathbb{T}$. Also, the general monothetic connected compact abelian group is a product $\mathbb{T}^m$ where card $m \leq c$. See Hewitt and Ross [7]. So what Proposition 3.6 proves is a category result for the case of τ being multiplication by a generator of a metric compact connected abelian group.

Because we have only a limited information about $O_n^\tau f$ for general discrete spectrum mappings, we cannot prove a category result even as good as the one in Proposition 3.6, let alone the result that we suggest in Remark 3.7. This blocks us at this time from answering the following question:

Question II: For a given discrete spectrum mapping τ, let $\mathcal{O}$ consist of the mean-zero functions in $L_2(X)$ such that for some K and N, we have $\|A_n^\tau f\|_2 \leq K O_n^\tau f$ for all $n \geq N$. Is $\mathcal{O}$ of first category? The issue is really this: is this result true for τ which has all of its eigenvalues being roots of unity.

Remark 3.9. Here are a couple of examples of what the issues seem to be that make it difficult to handle the case when there are eigenvalues that are roots of unity.

Suppose -1 is an eigenvalue of τ with eigenfunction χ. Then $\|A_n\chi\|_2$ is zero when n is even (and hence it is equal to $O_n^\tau\chi$), and it is $\frac{1}{n}$ if n is odd. But also it is not hard to see that, when n is odd, $O_n^\tau\chi$ is also $\frac{1}{n}$. So $\|A_n\chi\|_2 = O_n^\tau\chi$ for all n. On the other hand, if $i = \sqrt{-1}$ is an eigenvalue with eigenfunction χ, then $O_2^\tau\chi = 0$, because we can choose $m_1 = 0$, $m_2 = 2$. However, $A_2\chi = (i-1)\chi$, so $\|A_2\chi\|_2 > O_2^\tau\chi$. Examples of both of these types abound for other cases of roots of unity as eigenvalues and other choices of n.

4 Weak Mixing Case

Suppose first that τ is strongly mixing and $f \in L_2(X)$ is mean-zero. Consider choices $m_1, \dots, m_n$ that are distinct and such that $|m_i - m_j|$ is large when $i \neq j$. We know that $\|\sum_{k=1}^n f\circ\tau^{m_k}\|_2^2 = n\|f\|_2^2 + \sum_{i\neq j}\langle f, f\circ\tau^{m_i-m_j}\rangle$. Since τ is strongly mixing, we know that $\widehat{\nu_f^\tau}(m) = \langle f, f\circ\tau^m\rangle$ goes to zero as m goes to ∞. Hence, for fixed n,

$$\left|\sum_{i\neq j}\langle f, f\circ\tau^{m_i-m_j}\rangle\right| \leq \sum_{i\neq j}|\langle f, f\circ\tau^{m_i-m_j}\rangle|$$

will tend to zero as $m_i - m_j$ tends to ∞. So, $O_n^\tau f \leq \|f\|_2/\sqrt{n}$. Consider what happens in the special case that $f\circ\tau^k$ is an IID sequence. Then $\|\sum_{k=1}^n f\circ\tau^{m_k}\|_2^2$ is always $n\|f\|_2^2$. Hence, $O_n^\tau f = \|f\|_2/\sqrt{n}$. So it is not unreasonable to ask the following:

Question III: If τ is strongly mixing, is there a constant c such that $O_n^\tau f \geq c\|f\|_2/\sqrt{n}$ for all $n \geq 1$?

This question is related to the following for which an affirmative answer to either would resolve the issue:

Questions IV: If τ is strongly mixing, can we have $O_n^\tau f = 0$? More specifically, can we have $m_1, \dots, m_n$ such that $\sum_{k=1}^n f\circ\tau^{m_k} = 0$?

Remark 4.1. In Section 6, we answer these questions. But for now let us summarize the facts. The answers to Questions IV are negative. The answer to Question III, for a fixed n, is negative if c is meant to be independent of f. However, it is the case that if τ is strongly mixing, then at least $O_n^\tau f > 0$ for all non zero f.

When τ is just weakly mixing, then we only know that the spectral measures ν_f^τ have Fourier transforms that are mean-zero weakly almost periodic functions. In particular, we know that $\frac{1}{2n+1}\sum_{k=-n}^n|\langle f, f\circ\tau^k\rangle| \to 0$ as $n\to\infty$. This means that $|\widehat{\nu_f^\tau}(n)| = |\langle f, f\circ\tau^n\rangle| \to 0$ along a sequence $\mathbf{N}$ of density 1. It is not hard to see that therefore for all $\epsilon > 0$, we can inductively choose $m_1, \dots, m_n$ distinct such that for all $i \neq j$, $|\langle f, f\circ\tau^{m_i-m_j}\rangle| \leq \epsilon$. It follows that

$$\left\| \frac{1}{n} \sum_{k=1}^{n} f \circ \tau^{m_k} \right\|_2^2 \le \frac{1}{n} \|f\|_2^2 + \frac{1}{n^2} \sum_{i \ne j} |\langle f, f \circ \tau^{m_i - m_j} \rangle| \le \frac{\|f\|_2^2}{n} + \frac{n(n-1)\epsilon}{n^2}.$$

Since ϵ is arbitrary, this gives the following.

Proposition 4.2. *For τ weakly mixing and a mean-zero $f \in L_2(X)$, we have $O_n^\tau f \le \frac{\|f\|_2}{\sqrt{n}}$.*

Remark 4.3. Remark 2.5 showed that in a global sense this result is best possible even for strongly mixing transformations. However, for individual functions, this may not be best possible. For example, there are weakly mixing τ and functions f such that $O_2^\tau f = 0$. See the construction in Proposition 6.8. The result is that we do not know how to characterize, for an individual function, when the estimate in Proposition 4.2 is best possible.

Remark 4.4. There is always a dense class of functions $\mathcal{D}$, in the mean-zero functions in $L_2(X)$, such that for all $f \in \mathcal{D}$, there exists some K depending on f such that $\|A_n^\tau f\|_2 \le (K/n)\|f\|_2$ for all n. There are two different ways to see this. One approach is to just take coboundaries $f = g - g \circ \tau$ where $g \in L_2(X)$. Then with τ being ergodic, these are dense in the mean-zero functions and satisfy the norm estimate. Another approach is to consider the functions f such that for some $\delta > 0$, the spectral measure $\nu_f^\tau(A_\delta) = 0$ where A_δ is an arc of length 2δ symmetric around 1 in $\mathbb{T}$. These are also dense in the mean-zero functions. Now

$$\begin{aligned}
\|A_n^\tau f\|_2^2 &= \int_{\mathbb{T}} \left| \frac{1}{n} \sum_{k=1}^{n} \gamma^k \right|^2 d\nu_f^\tau(\gamma) \\
&\le \int_{\mathbb{T}\setminus A_\delta} \left| \frac{1}{n} \sum_{k=1}^{n} \gamma^k \right|^2 d\nu_f^\tau(\gamma) + \int_{A_\delta} \left| \frac{1}{n} \sum_{k=1}^{n} \gamma^k \right|^2 d\nu_f^\tau(\gamma) \\
&\le \frac{4}{n^2 |1 - \exp(i\delta)|^2} \nu_f^\tau(\mathbb{T}) + \nu_f^\tau(A_\delta) \\
&= \frac{8}{n^2 \delta^2} \|f\|_2^2 + \nu_f^\tau(A_\delta).
\end{aligned}$$

So, when $\nu_f^\tau(A_\delta)$ is zero, we have $\|A_n^\tau f\|_2 \le (K/n)\|f\|_2$ with $K = \sqrt{8}/\delta$.

We will need here a well-known lemma on norms.

Lemma 4.5. *For any ergodic map τ and $n \ge 1$, the operator norm of A_n^τ on the mean-zero functions in $L_2(X)$ is 1.*

Proof. The claim is that there is no constant $\delta < 1$ such that $\|A_n^\tau f\|_2 \le \delta\|f\|_2$ for all mean-zero functions in $L_2(X)$. Indeed, by Rokhlin's Lemma, for any N and

$\epsilon > 0$, we can choose a set E such that $\tau^i E, i = 1, \ldots, N$ are pairwise disjoint and $p(X \setminus \bigcup_{i=1}^N \tau^i E) < \epsilon$. Then write E as a disjoint union of E_1 and E_2, of the same measure, and let $f = \sum_{i=1}^N 1_{\tau^i E_1} - 1_{\tau^i E_2}$. Then f is mean-zero and $|f| = 1$ on $\bigcup_{i=1}^N \tau^i E$. Given $\delta > 0$ and $n \leq N$, by increasing N and decreasing ϵ if necessary, we would have $\|f - f \circ \tau^i\|_2 \leq \delta$ for all $i = 1, \ldots, n$, and $1 - \delta \leq \|f\|_2 \leq 1$. This shows $1 - \delta \leq \|A_n^\tau f\|_2 \leq 1$. Hence, letting $\delta \to 0$, we see that the norm of A_n^τ on the mean-zero functions in $L_2(X)$ is one. □

Remark 4.6. If T is a mean ergodic contraction on a Banach space E, but T is not uniformly ergodic, then $\|A_n^T\| = 1$ on $Y = \overline{(I - T)E}$. Since Koopman operators are not uniformly ergodic, this extends Lemma 4.5 to a more general class of operators.

Now we can use our results to prove a category result for weakly mixing mappings like Proposition 3.6.

Proposition 4.7. *Assume τ is weakly mixing. Fix a subsequence $\mathbf{N}$ of $\mathbb{N}$. Let $\mathcal{O}(\mathbf{N})$ consist of all mean-zero functions such that for some K and N, we have $\|A_n^\tau f\|_2 \leq K O_n^\tau f$ for all $n \geq N, n \in \mathbf{N}$. Then $\mathcal{O}(\mathbf{N})$ is a set of first category.*

Proof. Again, we are considering $\mathcal{O}(\mathbf{N}) = \bigcup_{K=1}^\infty \bigcup_{N=1}^\infty \bigcap_{n \geq N, n \in \mathbf{N}} \mathcal{O}_{K,n}$ where $\mathcal{O}_{K,n} = \{f : \int f\,dp = 0, \|A_n^\tau f\|_2 \leq K\, O_n^\tau f\}$. We saw in Proposition 3.6 that each $\mathcal{O}_{K,n}$ is closed and so $\bigcap_{n \geq N, n \in \mathbf{N}} \mathcal{O}_{K,n}$ is closed. We claim that $\bigcap_{n \geq N, n \in \mathbf{N}} \mathcal{O}_{K,n}$ has no interior, and so $\mathcal{O}(\mathbf{N})$ is a set of first category.

Suppose $\bigcap_{n \geq N, n \in \mathbf{N}} \mathcal{O}_{K,n}$ contains interior. Then there exists f_0 and $\epsilon_0 < 1$ such that for all mean-zero g, $\|g\|_2 \leq 1$, we have $f_0 + \epsilon_0 g$ in this set. Choose $n \geq N, n \in \mathbf{N}$. By Proposition 4.2, we have

$$\begin{aligned}
\|A_n^\tau(\epsilon_0 g)\|_2 &\leq \|A_n^\tau f_0\|_2 + \|A_n^\tau(f_0 + \epsilon_0 g)\|_2 \\
&\leq K O_n^\tau f_0 + K O_n^\tau (f_0 + \epsilon_0 g) \\
&\leq \frac{K\|f_0\|_2}{\sqrt{n}} + \frac{K\|f_0 + \epsilon_0 g\|_2}{\sqrt{n}} \\
&\leq \frac{2K\|f_0\|_2}{\sqrt{n}} + \frac{K\epsilon_0}{\sqrt{n}} \\
&\leq \epsilon_0^2
\end{aligned}$$

for some particular n which is sufficiently large. But then for this value of n, $\|A_n^\tau g\|_2 \leq \epsilon_0$ for all mean-zero g with $\|g\|_2 \leq 1$. It follows that the operator norm of A_n^τ on the mean-zero functions is strictly less than 1. By Lemma 4.5, this cannot be the case. □

Remark 4.8. In Proposition 4.7 we cannot use an arbitrary (K_n) as we did in Proposition 3.6. This is because we are relying on Proposition 4.2, which does not give as strong a result as is available for some discrete spectrum dynamical systems for which there is a dense class of functions f with $O_n^\tau f = 0$ eventually.

It is easy to see that there is nothing particular about the use of the usual averages in this result. Indeed, the following is true.

Proposition 4.9. *Assume τ is weakly mixing. Let $\mathcal{A}_n$ be any sequence of probability measures on $\mathbb{Z}$. Fix a subsequence $\mathbf{N}$ of $\mathbb{N}$. Let $\mathcal{O}(\mathbf{N})$ consist of all mean-zero functions such that for some K and N, we have $\|\mathcal{A}_n^\tau f\|_2 \leq KO_n^\tau f$ for all $n \geq N, n \in \mathbf{N}$. Then $\mathcal{O}(\mathbf{N})$ is a set of first category.*

Proof. In this case, we are not even assuming that $\mathcal{A}_n$ consists of a convex combination of n points. But it is still the case that the norm of $\mathcal{A}_n^\tau$ on the mean-zero functions in $L_2(X)$ is one. This was the critical point needed to conclude the proof. □

Remark 4.10. It may be possible to extend this to other dynamical systems, e.g., the case where τ is an ergodic rotation of the circle. At least the techniques that appear later in Section 7 can be used to show that Proposition 4.9 holds for a residual set of transformations too.

5 General Case

We would like to put together the category results from Section 3 and Section 4. The idea is to use the standard decomposition of $L_2(X)$ into the orthogonal spaces $L_{2,c}(X)$, which is spanned by the eigenfunctions for τ, and $L_{2,w}(X)$, the orthogonal complement of $L_{2,c}(X)$. Given $f \in L_2(X)$, we can uniquely write $f = f_1 + f_2$, with $f_1 \in L_{2,c}(X)$ and $f_2 \in L_{2,w}(X)$. The property that f_2 has is that $\widehat{\nu_{f_2}^\tau}$ is a mean-zero weakly almost periodic function on $\mathbb{Z}$. Also, the Koopman operator for τ restricted to $L_{2,c}(X)$ is isomorphic to the Koopman operator for multiplication by a generator g of some metric compact abelian group such that $\langle g\rangle$, the subgroup generated by g, is dense in G, i.e., G is monothetic with a generator g. See Bergelson and Rosenblatt [1].

The difficulty is that it is not clear how to choose values of $m_1, \ldots, m_n$ so that we simultaneously decrease the correlations $\langle f_2, f_2 \circ \tau^{m_i - m_j}\rangle$ and also have $\frac{1}{n}|\sum_{k=1}^n \alpha^{m_k}|$ small for sufficiently many choices of eigenvalues for τ. This is even an issue when τ restricted to $L_{2,c}$ is isomorphic to an ergodic rotation of $\mathbb{T}$. The result is that we can only estimate $O_n^\tau f$ for special values of large n instead of arbitrary values.

The question is this: Can we at least get the same type of result as the one conjectured for discrete spectrum mappings? See the end of Section 3.

Question V: Given an ergodic τ, let $\mathcal{O}$ consist of the mean-zero functions in $L_2(X)$ such that for some K and N, we have $\|A_n^\tau f\|_2 \leq KO_n^\tau f$ for all $n \geq N$. Is $\mathcal{O}$ a set of first category?

Remark 5.1. We have proved a somewhat better result than this for the case of discrete spectrum mappings with no roots of unity as eigenvalues, and for weakly mixing mappings. Below we indicate how to prove this in the mixed spectrum case. This leaves the answer to this question open only for discrete spectrum mappings which have some eigenvalues that are roots of unity.

We will need a lemma that generalizes Lemma 4.5 in cases where the ergodic mapping τ is not a discrete spectrum map.

Lemma 5.2. *Suppose τ does not have discrete spectrum. Then for any $n \geq 1$, the operator A_n^τ restricted to the functions in $L_{2,w}(X)$ is norm one.*

Proof. Take a non zero $h \in L_{2,w}(X)$. Take an eigenvalue α and eigenfunction χ for τ. Then $\langle(\chi h)\circ\tau^k, \chi h\rangle = \alpha^k\langle\chi(h\circ\tau^k), \chi h\rangle = \alpha^k\langle h\circ\tau^k, h\rangle$. Hence, $|\langle(\chi h)\circ\tau^k, \chi h\rangle| = |\langle h\circ\tau^k, h\rangle|$ for all k. This shows $\chi h \in L_{2,w}(X)$. Indeed, this also shows that the spectral measure $\nu_{\chi h}^\tau$ is the rotation $\nu_h^\tau * \delta_\alpha$ of ν_h^τ. Now fix a small arc I about $1 \in \mathbb{T}$. By rotating ν_h^τ, and then restricting it to I, we can obtain a non zero, continuous measure $\nu \ll \nu_{\chi h}^\tau$ which is supported in I. But there exists some non zero $h_I \in L_{2,w}(X)$ such that $\nu = \nu_{h_I}^\tau$. Now we have $\|h_I \circ \tau - h_I\|_2^2 = \int_I |\gamma - 1|^2\, d\nu_{h_I}^\tau(\gamma)$. However, for any $\epsilon > 0$, if I is sufficiently small, then $\int_I |\gamma-1|^2\, d\nu_{h_I}^\tau(\gamma) \leq \epsilon^2\|\nu_{h_I}^\tau\|_1$. That is, $\|h_I\circ\tau - h_I\|_2^2 \leq \epsilon^2\|h_I\|_2^2$. Hence, for any $\delta > 0$, by decreasing ϵ as needed, this gives a function $H = h_I/\|h_I\|_2$, which is in $L_{2,w}(X)$ such that $\|A_n^\tau H\| \geq 1-\delta$. Thus, for any n, the norm of A_n^τ on $L_{2,w}(X)$ is one. ☐

Now let us first prove the following special case. We will use this to prove a more general result.

Proposition 5.3. *Assume τ ergodic and has no roots of unity as eigenvalues, but is not a discrete spectrum mapping. Let $\mathcal{O}(\mathbb{N})$ consist of all mean-zero functions such that for some K and N, we have $\|A_n^\tau f\|_2 \leq KO_n^\tau f$ for all $n \geq N$. Then $\mathcal{O}(\mathbb{N})$ is a set of first category.*

Proof. Again, we are considering $\mathcal{O}(\mathbb{N}) = \bigcup_{K=1}^\infty \bigcup_{N=1}^\infty \bigcap_{n\geq N} \mathcal{O}_{K,n}$ where $\mathcal{O}_{K,n} = \{f : \int f\, dp = 0, \|A_n^\tau f\|_2 \leq K\, O_n^\tau f\}$. We saw in Proposition 3.6 that each $\mathcal{O}_{K,n}$ is closed and so $\bigcap_{n\geq N}\mathcal{O}_{K,n}$ is closed. We claim that $\bigcap_{n\geq N}\mathcal{O}_{K,n}$ has no interior, and so $\mathcal{O}(\mathbb{N})$ is a set of first category.

Suppose $\bigcap_{n\geq N}\mathcal{O}_{K,n}$ contains interior. Then there is a finite linear combination of characters f_1, some $f_2 \in L_{2,w}(X)$, and $0 < \epsilon_0 < 1$, such that for all $g \in L_2(X)$, $\|g\|_2 \leq 1$, we have $f_1 + f_2 + \epsilon_0 g \in \bigcap_{n\geq N}\mathcal{O}_{K,n}$. We can write $g = g_1 + g_2$ with $g_1 \in L_{2,c}(X)$, $g_2 \in L_{2,w}(X)$, and $\|g_1\|_2^2 + \|g_2\|_2^2 = \|g\|_2^2$. Suppose in fact that $g_1 = 0$. Then, take $n \geq N$ and any $\epsilon > 0$. For large enough n, and suitable probability measures μ_{n^2} and ν_{n^2} (which will actually each be products of measures using only n Dirac masses), we would have

$$\begin{aligned}
\|A_{n^2}^\tau(\epsilon_0 g)\|_2 &\le \|A_{n^2}^\tau(f_1+f_2+\epsilon_0 g_2)\|_2 + \|A_{n^2}^\tau(f_1+f_2)\|_2 \\
&\le KO_n^\tau(f_1+f_2+\epsilon_0 g_2) + KO_n^\tau(f_1+f_2) \\
&\le K\|\mu_{n^2}^\tau(f_1+f_2+\epsilon_0 g_2)\|_2 + K\|\nu_{n^2}^\tau(f_1+f_2)\|_2 \\
&\le K\left(\epsilon + \frac{\|f_2+\epsilon_0 g_2\|_2}{\sqrt{n}}\right) + K\left(\epsilon + \frac{\|f_2\|_2}{\sqrt{n}}\right) \\
&\le 2K\epsilon + 2K\frac{\|f_2\|_2}{\sqrt{n}} + K\frac{\epsilon_0}{\sqrt{n}}.
\end{aligned}$$

Hence, for sufficiently large n and small ϵ, we would have $\|A_{n^2}^\tau(\epsilon_0 g_2)\|_2 \le \epsilon_0^2$. But this means that $A_{n^2}^\tau$ is norm less than one when restricted to $L_{2,w}$. By Lemma 5.2, this cannot hold. □

We can actually use this method to give a general mixed mapping result of the same type when τ has some non zero weakly mixing functions. Let $\mathcal{E}$ be the eigenvalues of τ. There are two cases. In the first case, either $\mathcal{E}$ generates a finite group or an infinite one that is automatically dense in $\mathbb{T}$. In the second case, we can use the argument in Proposition 5.3 if we always work with measures μ_n, ones having n point masses. Then the observation in Proposition 3.5 allows us to again have $\|\mu_n^\tau f_1\|_2$ small no matter what types of eigenvalues arise in considering f_1. In the first case, choose some fixed N such that $\alpha^N = 1$ for all $\alpha \in \mathcal{E}$. Then, we can again use the argument above, but now we work with measures μ_{N^e}, ones having N^e point masses where e is the cardinality of $\mathcal{E}$. Then we can actually guarantee that $\mu_{N^e}^\tau f_1 = 0$ for all $f_1 \in L_{2,c}(X)$, which is now a finite dimensional space. We would need to again give an argument for the norm of A_n^τ being one on $L_{2,w}(X)$, for any n. But this follows since the spectrum of the Koopman operator given by τ is all of $\mathbb{T}$. Then since $\mathcal{E}$ is finite, there must be functions $h \in L_{2,w}(X)$ whose spectral measure has support close to 1. Then the argument in Proposition 5.3 can be used again.

The result of this discussion is the following generalization of Proposition 5.3.

Proposition 5.4. *Assume τ ergodic and not a discrete spectrum mapping. Let $\mathcal{O}(\mathbb{N})$ consist of all mean-zero functions such that for some K and N, we have $\|A_n^\tau f\|_2 \le KO_n^\tau f$ for all $n \ge N$. Then $\mathcal{O}(\mathbb{N})$ is a set of first category.*

Vitaly Bergelson asked the following question: fix a non zero, mean-zero $f \in L_2(X)$ and consider varying the choice of the transformation. Then is the typical behavior for τ that one has $O_n^\tau f < \epsilon\|A_n^\tau f\|_2$ for many (or most) values of n? This is in some sense the dual version of what we have considered in the category results above, where τ was fixed and the function f varied. This issue is discussed more in Section 7, but for now we can prove fairly simply a result of this type. Let $\mathcal{T}$ denote the measure-preserving transformations in the usual weak topology.

Proposition 5.5. *There is a residual set $\mathcal{R}$ in the mean-zero functions in $L_2(X)$ such that for every $f \in \mathcal{R}$, there is a residual set $\mathcal{S}$ in $\mathcal{T}$ such that for all K and $\tau \in \mathcal{S}$, we have $O_n^\tau f < \frac{1}{K}\|A_n^\tau f\|_2$ for infinitely many values of n.*

Proof. The weak mixing transformations are a residual class in $\mathcal{T}$. For each weak mixing τ, there is a residual class of functions such that for each K, we have $O_n^\tau f < \frac{1}{K}\|A_n^\tau f\|_2$ for infinitely many values of n. Consider the pairs $(f,\tau) \in L_2(X) \times \mathcal{T}$ such that for each K we have $O_n^\tau f < \frac{1}{K}\|A_n^\tau f\|_2$ for infinitely many values of n. This is a Borel set and hence a set with the property of Baire. So we can use the Kuratowski–Ulam Theorem, see Theorem 15.4 in Oxtoby [10]. This theorem allows us to reverse the roles of f and τ, giving our proposition. □

Remark 5.6. It is not clear if this result can be proved for every mean-zero $f \in L_2(X)$. Of course, using the argument above, this would be true if we could show that any f and K, there is a dense set of τ such that $O_n^\tau f < (1/K)\|A_n^\tau\|_2$ for infinitely many n.

6 Properties of O_n^τ

We know that we can have $\sum_{k=1}^n f \circ \tau^{m_k} = 0$. For example, let τ have an eigenvalue $\alpha \neq 1$ which is an n-th root of unity so that $\sum_{k=1}^n \alpha^k = 0$. Take a non zero eigenfunction f. Then $\sum_{k=1}^n f \circ \tau^k = \sum_{k=1}^n \alpha^k f = 0$. It is natural to ask if this is the only way that this can happen?

The answer is negative. We thank Bruce Reznick for providing the following construction.

Proposition 6.1. *There is $\alpha \in \mathbb{T}$ which is not a root of unity such that for suitable $m_1, \ldots, m_n \in \mathbb{N}$, we have $\sum_{k=1}^n \alpha^{m_k} = 0$.*

Proof. The first fact is that four of the roots of the sixth degree polynomial $P(x) = x^6 + x^5 + x^3 + x + 1$ lie on the unit circle. Since they are algebraic numbers of degree less than or equal to 6, if they were roots of unity, then they would have a cyclotomic polynomial as their minimal polynomial, and it would have to divide P. There are only a few cyclotomic polynomials of degree less than or equal to 6 and none of them work here.

To derive this example, consider the cubic $p(T) = T^3 - T^2 - 3T + 1$. Since $p(0) = 1$ and $p(1) = -2$, p has at least one root between 0 and 1. In fact, a computation shows that the roots are approximately $-1.48\ldots$, $311\ldots$, and $2.17\ldots$. Let these roots be a, b, and c, so that $a + b + c = 1$, $ab + ac + bc = -3$, and $abc = -1$. Now consider the polynomial $(x^2 + ax + 1)(x^2 + bx + 1)(x^2 + cx + 1) = (x^6 + 1) + (a + b + c)(x^5 + x) + (3 + ab + ac + bc)(x^4 + x^2) + (abc + 2(a + b + c))x^3$ which is $(x^6 + 1) + (x^5 + x) + x^3$. This the example above. Since a and b are in the interval $[-2, 2]$, the roots of $(x^2 + ax + 1)$ and $(x^2 + bx + 1)$ lie on the unit circle. □

Remark 6.2. Of course, if α is not an algebraic number, then for all finite (positive) measures μ, we would have $\widehat{\mu}(\alpha) \neq 0$.

Remark 6.3. Given α in Proposition 6.1 and $j \in \mathbb{Z}$, does there exist a Cesàro average μ_n, or at least a positive finite measure μ_n such that $\widehat{\mu}_n(\alpha^j) = 0$? If so, then we can build by products a sequence of positive Cesàro averages (or positive finite measures) ω_n such that for all J, there exists N_J such that $\widehat{\omega}_n(\alpha^j) = 0$ for all $|j| \leq J$ when $n \geq N_J$. But then for τ being rotated by α on $\mathbb{T}$, for any linear combination $\sum_{|j|\leq J} c_j \chi_j$, where $\chi_j(\gamma) = \gamma^j$, we would have $\|\omega_n^\tau f\|_2 = 0$ once n is sufficiently large. Thus, for weighted ergodic averages ω_n^τ, we can have the averages giving optimal norm approximation on a dense subspace as the index n tends to ∞.

But it is the case that only mappings with some discrete spectrum can have $\sum_{k=1}^n f \circ \tau^{m_k} = 0$ in a non trivial manner. More generally, we have the following.

Proposition 6.4. *Suppose that for some non zero, mean-zero $f \in L_2(X)$, there are distinct $m_1, \dots, m_n$ and non zero $c_1, \dots, c_n \in \mathbb{C}$ such that $\sum_{k=1}^n c_k f \circ \tau^{m_k} = 0$. Then the linear span of $\{f \circ \tau^k : k \in \mathbb{Z}\}$ is finite dimensional and τ has non trivial eigenvalues. In particular, τ cannot be weakly mixing.*

Proof. By composing with a power of τ, we may suppose that for some n, possibly different than the original value, we have $\sum_{k=1}^n c_k f \circ \tau^k = 0$ where c_1 and c_n are not zero, but the other coefficients might be zero. Then consider the linear span S of $\{f \circ \tau^k : k = 1, \dots, n-1\}$. For each $k = 1, \dots, n-1$, we have $(f \circ \tau^k) \circ \tau$ in this span. Hence, $S \circ \tau \subset S$. But S is finite-dimensional, so $S \circ \tau = S$ and S is τ invariant. Because the Koopman operator $Uf = f \circ \tau$ is unitary and S is finite-dimensional, this means that τ must have non trivial eigenvalues $\alpha \in \mathbb{T}$ with eigenfunctions in S. □

Remark 6.5. This result says that for weakly mixing transformations τ, the functions $\{f \circ \tau^k : k \in \mathbb{Z}\}$ are linearly independent. Characterizing when there can be linear dependence instead becomes then an issue of knowing what algebraic complex numbers $\alpha \in \mathbb{T}$ are actually eigenvalues of transformations τ.

Proposition 6.4 actually allows us to show that if τ is strongly mixing, then $O_n^\tau f$ is never zero. The idea of this argument can be understood better if we take $m_i - m_j$ very large for $i \neq j$. Then the functions $f \circ \tau^{m_i}$ are essentially uncorrelated and $\|\frac{1}{n}\sum_{k=1}^n f \circ \tau^{m_k}\|_2$ gets close to $\|f\|_2/\sqrt{n}$. But then at the other extreme, if the values of $m_i - m_j$

stay bounded, then Proposition 6.4 shows that $\|\frac{1}{n}\sum_{k=1}^{n} f\circ\tau^{m_k}\|_2$ must be bounded away from zero.

Proposition 6.6. *If τ is strongly mixing, then $O_n^\tau f$ is never zero.*

Proof. We have to formalize the argument above. Fix n and take $m_1(s),\dots,m_n(s)$ such that $\|\frac{1}{n}\sum_{k=1}^{n} f\circ\tau^{m_k(s)}\|_2$ converges to $O_n^\tau f$ as $s\to\infty$. We can pass to a subsequence if necessary and assume there are blocks of the indices $\{1,\dots,n\}$, say $B_1(s),\dots,B_R(s)$, such that for $|m_i(s)-m_j(s)|$ tends to ∞ as $s\to\infty$ when i and j are in different blocks, but $|m_i(s)-m_j(s)|$ is bounded when i,j are in the same block. Then for large s, we would have $\|\frac{1}{n}\sum_{k=1}^{n} f\circ\tau^{m_k(s)}\|_2^2$ very close to being $\sum_{r=1}^{R}\|\frac{1}{n}\sum_{k\in B_r} f\circ\tau^{m_k(s)}\|_2^2$. If this expression is going to zero, then it must be that the terms $\|\frac{1}{n}\sum_{k\in B_r} f\circ\tau^{m_k(s)}\|_2^2$ are going to zero. But the boundedness of the gaps on such blocks means that there would be some distinct powers p_l such that $\|\sum_l f\circ\tau^{p_l}\|_2 = 0$. Since τ is strongly mixing and has no non trivial eigenfunctions, Proposition 6.4 shows this cannot happen. □

Remark 6.7.

(a) This result can be rephrased as follows. Given τ, which is strongly mixing, and a mean-zero $f\in L_2(X)$, there is a constant $c>0$ such that for all $m_1,\dots,m_n\in\mathbb{Z}$, we have $\|\frac{1}{n}\sum_{k=1}^{n} f\circ\tau^{m_k}\|_2\ge c$. The constant c depends on n and f. Also, more generally, we can use this argument to show that if ν is a positive probability measure on $\mathbb{T}$ whose Fourier transform vanishes at infinity, then there is a constant $c>0$ such that for all $m_1,\dots,m_n\in\mathbb{Z}$, we have $\int_{\mathbb{T}}|\frac{1}{n}\sum_{k=1}^{n}\gamma^{m_k}|^2\,d\nu(\gamma)\ge c$.

(b) The inequality in Remark 6.7 (a) cannot be made homogeneous. That is, if $n\ge 2$, we cannot have a constant $c>0$ such that for all mean-zero f in $L_2(X)$, we have $O_n^\tau f\ge c\|f\|_2$. Indeed, if this held for some ergodic mapping (let alone strongly mixing mapping) τ_0, then we would also have $\|A_n^{\tau_0}f\|_2\ge c\|f\|_2$ for all mean-zero f in $L_2(X)$. But then, $\|A_n^{\sigma\circ\tau_0\circ\sigma^{-1}}f\|_2\ge c\|f\|_2$. So letting σ vary and using weak approximation, we would have for any invertible measure-preserving transformation τ, $\|A_n^\tau f\|_2\ge c\|f\|_2$. But this is impossible because, using an appropriate τ and an eigenvector f, for any $n\ge 2$ there is a non zero, mean-zero f and some τ such that $A_n^\tau f=0$.

But Proposition 6.6 is not true for weakly mixing mappings. We really needed to use the fact the $\langle f, f\circ\tau^m\rangle$ tends to zero as $n\to\infty$, not just that this is true along a sequence of density one. Indeed, we have the following anti-rigidity construction.

Proposition 6.8. *There is a weakly mixing transformation τ, a function $f\in L_2(X)$, and a sequence $(n_m : m\ge 1)$ such that $f\circ\tau^{n_m}\to -f$ in L_2-norm as $m\to\infty$. Hence, $O_2^\tau f = 0$.*

Proof. The idea of the construction is that we first build an infinite, compact set $K \subset \mathbb{T}$ and (n_m) such that for all $\gamma \in K$, $\gamma^{n_m} \to -1$ as $m \to \infty$. Once this is done, take any continuous probability measure ν on K. Then the Gaussian measure space construction gives us a weakly mixing τ on a standard probability space (X, β, p), and a mean-zero function $f \in L_2(X)$, such that $\widehat{\nu}(k) = \langle f, f \circ \tau^k \rangle$ for all $k \in \mathbb{Z}$. But $\widehat{\nu}(n_m) \to -1$ as $m \to \infty$. So $\|f + f \circ \tau^{n_m}\|_2^2 = 2\|f\|_2^2 + \widehat{\nu}(n_m) + \widehat{\nu}(-n_m) = 2 + \widehat{\nu}(n_m) + \widehat{\nu}(-n_m) \to 0$ as $m \to \infty$. Hence, $O_2^\tau f = 0$.

To construct K, use a standard method but build in the necessary anti-rigidity. We work with successive choices of even numbers p_i which are increasing to ∞. We start by choosing two small disjoint closed arcs $A(1)$ and $A(2)$ around two p_1-th roots such that $\alpha^{p_1/2}$ is close to -1 for all α in both arcs. Then we take p_2 much larger than p_1, as needed, so that we can form two small disjoint closed arcs $A(i_1, i_2)$ within each $A(i_1)$ on which we have $\alpha^{p_2/2}$ even closer to -1 for all α in the four constituent arcs formed at this second step of the construction. Continuing in this fashion to construct arcs $A(i_1, \ldots, i_d)$, we can form a compact set $K = \bigcap_d \bigcup_{(i_1,\ldots,i_d)} A(i_1, \ldots, i_d)$ with the desired property. □

Remark 6.9. It should be possible to give all types of actually different examples along the lines of the example in Proposition 6.8, but one such example may suffice for our purposes here.

Here is another well-known limit estimate which raises an interesting question.

Question VI: Take $f = g - g \circ \tau$. Then $\|A_n^\tau f\|_2 \le 2\|g\|_2/n$. This gives a good rate estimate. But what is $O_n^\tau f$ in this case?

7 Oscillation and generic behavior for transformations

It may be possible to have companion results to the ones in Sections 3, 4, and 5 that redeem the use of $\|A_n^\tau f\|_2$ instead of $O_n^\tau f$, at least infinitely often. The idea is to show the opposite of what we have obtained already for $\|A_n^\tau f\|_2$ and $O_n^\tau f$. We have seen that for most τ, possibly for all ergodic τ, the generic mean-zero function $f \in L_2(X)$ has $O_n^\tau f < \frac{1}{K}\|A_n^\tau f\|_2$ infinitely often. But this inequality may not actually hold eventually, even in the generic case; at least we have not been able to prove this even for a particular transformation. So it is possible that there are functions, in fact possibly the generic function, such that for all $\delta > 0$, $O_n^\tau f \ge (1 - \delta)\|A_n^\tau f\|_2$ infinitely often. At this time, it is not clear when this happens, but at least we know it can happen. See the generic results at the end of this section. Also, there may be such a result with f fixed where one replaces $1 - \delta$ by some constant $D < 1$ to ask for a less dramatic comparison of $O_n^\tau f$ and $\|A_n^\tau f\|_2$. We also do not know if there is an example of a transformation with this property where the stronger property fails.

However, if this type of result did hold generically for a given τ, then we would be showing that there must be an oscillation between $\|A_n^\tau f\|_2$, being a bad estimate for $O_n^\tau f$, and $\|A_n^\tau f\|_2$, being a good estimate for $O_n^\tau f$. While we are not sure when this happens, the idea does suggest that we should look at the issue of the oscillation behavior of $\|A_n^\tau f\|_2$ and $O_n^\tau f$ by themselves.

7.1 Oscillation of $O_n^\tau f$ and $\|A_n^\tau f\|_2$

First, the results that we have obtained do show that there is no saturation limit for $O_n^\tau f$. It can be eventually zero in non trivial cases; see Proposition 3.1. But there is a saturation limit for $\|A_n^\tau f\|_2$, i.e., a rate below which one cannot go without the averages becoming trivial. First, note that for any given ergodic transformation τ, if $f = F - F \circ \tau$, with $F \in L_2(X)$, then $O_n^\tau f \leq \|A_n^\tau f\|_2 \leq 2\|F\|_2/n$ for all $n \geq 1$. In addition, these functions f, coboundaries with respect to τ, form a L_2-norm dense subspace of the mean-zero functions. At the same time, we have the following well-known fact. See Butzer and Westphal [4].

Proposition 7.1. *Suppose $f \in L_2(X)$ is mean-zero and $\|A_n^\tau f\|_2 = o(1/n)$. Then $f = 0$.*

Proof. This rate estimate says that $S_n^\tau f = \sum_{k=1}^n f \circ \tau^k$ has $\|S_n^\tau f\|_2 = o(1)$. But then $f \circ \tau - f \circ \tau^{n+1} = S_n^\tau f - (S_n^\tau f) \circ \tau$ shows that $\|f \circ \tau - f \circ \tau^{n+1}\|_2 = o(1)$ too. It follows that $\|f \circ \tau - A_N^\tau f\|_2 = o(1)$ because $\|f \circ \tau - A_N^\tau f\|_2 = \|\frac{1}{N}\sum_{n=1}^N (f \circ \tau - f \circ \tau^n)\|_2 \leq \frac{1}{N}\sum_{n=1}^N \|f \circ \tau - f \circ \tau^n\|_2$. But $\|A_N^\tau f\|_2 \to 0$ as $N \to \infty$, i.e., $\|A_N^\tau f\|_2 = o(1)$. So we must have $\|f \circ \tau\|_2 = o(1)$, i.e., $f = 0$. □

Remark 7.2. This result can certainly be extended. For example, it is also true for one-to-one power-bounded operators on $L_2(X)$, and if we take our sums $S_n^\tau f$ to be over $0 \leq k \leq n-1$, then the one-to-one hypothesis is not needed.

Remark 7.3. Is it possible to show there is a L_2-norm dense subspace of the mean-zero functions such that for every f in the subspace and every τ, we have $\|A_n^\tau f\|_2 = O(1/n)$? This would make it so that the functions demonstrating the saturation limit given by Proposition 7.1 could be chosen independently of the transformation. See Remark 7.5 and Remark 7.7 where a class with a rate property can be chosen independently of the transformation. The answer to this is negative; one can show that for a given non zero $f \in L_r(X)$, $1 \leq r \leq \infty$, the generic transformation τ does not have f as a coboundary with transfer function in $L_r(X)$ too. See Rosenblatt [12].

Even though there is no saturation limit for $O_n^\tau f$, there are constraints on how small it can be. For example, we saw in Proposition 2.7 that for a residual set of functions $O_n^\tau f$ is never zero. In addition, the method of proof of Proposition 2.7 gives the following pseudo-saturation statement.

Proposition 7.4. *Let $\epsilon(n) = o(1/n)$. Then there is a residual set $\mathcal{R}$ in the mean-zero functions in $L_2(X)$ such that for all $f \in \mathcal{R}$, we have $O_n^\tau f > \epsilon(n)$ infinitely often.*

Proof. There is no harm in assuming $\epsilon(n) > 0$ for all n. We show that $\bigcup_{N=1} \bigcap_{n\geq N} \{O_n^\tau f \leq \epsilon(n)\}$ is a first category set. First, notice that $\{O_n^\tau f \leq \epsilon(n)\}$ is closed in the L_2-norm topology. To see this, suppose (f_s) has $O_n^\tau f_s \leq \epsilon(n)$ for all s and $\|f - f_s\|_2 \to 0$ as $s \to \infty$. Then for $\epsilon > 0$, choose $\mu_n(s)$ such that $\|\mu_n(s)^\tau f_s\|_2 \leq \epsilon(n) + \epsilon$. Then $\|\mu_n(s)^\tau f\|_2 \leq \|f - f_s\|_2 + \epsilon(n) + \epsilon$. Hence, $O_n^\tau f \leq \|f - f_s\|_2 + \epsilon(n) + \epsilon$. Let $s \to \infty$, and then $\epsilon \to 0^+$, to show that $O_n^\tau f \leq \epsilon(n)$ too. So each $\{O_n^\tau f \leq \epsilon(n)\}$ and hence $\bigcap_{n\geq N}\{O_n^\tau f \leq \epsilon(n)\}$ are closed in the L_2-norm topology.

To finish the proof, we just need to show that $\bigcap_{n\geq N}\{O_n^\tau f \leq \epsilon(n)\}$ contains no interior. If it did contain interior, then following the same type of argument and notation as in the proof of Proposition 2.7, for $n \geq N$ and $\epsilon > 0$, we can choose μ_n and ν_n so that

$$\|(\mu_n * \nu_n)^\tau(\epsilon_0 g)\|_2 \leq 2(\epsilon(n) + \epsilon).$$

But then $O_{n^2}^\tau(\epsilon_0 g) \leq 2\epsilon(n)$. It follows that $O_{n^2}^\tau(g) \leq (2\epsilon(n)/\epsilon_0)$. Taking g as in the proof of Proposition 2.7, with $p(E)$ small and n large, would then give $\frac{1}{2n} \leq (\epsilon(n)/\epsilon_0)(2\|f_0\|_2 + \epsilon_0)$ for n as large as we like. But since $\epsilon(n) = o(1/n)$, this cannot hold for large enough values of n. □

Remark 7.5. For some functions and rates, one can always get a lower estimate that holds for all n. Take λ irrational and let $f = \lambda 1_E - 1_{E^c}$, where $p(E) = 1/(1+\lambda)$. Then f is mean-zero. Also, for any τ, and any representative measure $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{m_k}$, we have $\mu_n^\tau(f) \geq \frac{1}{n}\lambda_n$, where $\lambda_n = \min\{|r\lambda - (n-r)| : r = 0, \ldots, n\}$. Because λ is irrational, $\lambda_n > 0$. Let $\epsilon(n) = \frac{1}{n}\lambda_n$. Then $\epsilon(n) > 0$ and $\|\mu_n^\tau f\|_2 \geq \epsilon(n)$ for all τ and n. Hence, $O_n^\tau f \geq \epsilon(n)$ for all n.

On the other hand, at least for some dynamical systems, we cannot have $O_n^\tau f$ eventually larger than a non zero function $\epsilon(n)$, no matter how quickly $\epsilon(n)$ tends to zero, for a residual class of functions f. Here is a proof of this for discrete spectrum maps. Later in this section we also show this behavior occurs for the generic map also.

Proposition 7.6. *Suppose τ has discrete spectrum and $\epsilon(n) > 0$ for all n. Then there is a residual set $\mathcal{R}$ of mean-zero functions $f \in L_2(X)$ such that for all N, there exists $n \geq N$, such that $O_n^\tau f < \epsilon(n)$.*

Proof. Consider the set $\bigcup_{N=1}^{\infty}\bigcap_{n\geq N}\{O_n^\tau f \geq \epsilon(n)\}$. It is easy to show that each $\{O_n^\tau f \geq \epsilon(n)\}$ is an L_2-norm closed set. So to prove this result, we only need to show that $\bigcap_{n\geq N}\{O_n^\tau f \geq \epsilon(n)\}$ contains no interior. But by Proposition 3.5, there is a dense set of functions f such that $O_n^\tau f = 0$ for infinitely many n. So if $\bigcap_{n\geq N}\{O_n^\tau f \geq \epsilon(n)\}$ contained interior, there would be some $f \in \bigcap_{n\geq N}\{O_n^\tau f \geq \epsilon(n)\}$ which also had $O_n^\tau f = 0$ infinitely often. This is impossible because $\epsilon(n) > 0$ for all n. □

Remark 7.7. Remark 7.5 shows that an overestimate for $O_n^\tau f$, holding infinitely often, cannot hold for all functions if the rate $\epsilon(n)$ is chosen appropriately. So in contrast to Proposition 7.6 (see also Proposition 7.11), it would be worthwhile to know if examples such as in Remark 7.5 can be chosen to be dense in the mean-zero functions, perhaps depending on the choice of τ. Indeed, the following construction works. Let $\epsilon(n)$ decrease to zero exponentially, say $\epsilon(n) = 1/2^n$. Consider a function of the form $f = \sum_{l=1}^{L}\alpha_l 1_{E_l} - 1_F$ where the E_l are pairwise disjoint with $p(E_l) = \gamma$ for $l = 1,\dots,L$, F is the complement of $\bigcup_{l=1}^{L}E_l$, and $\gamma\sum_{l=1}^{L}\alpha_l - (1 - L\gamma) = 0$. Then f is mean-zero. Such mean-zero functions are dense in the mean-zero functions in $L_2(X)$. Indeed, this remains true if we also restrict ourselves to choices of $\{\alpha_l : l = 0,\dots,L\}$ which consists of algebraic integers that are jointly rationally independent, where we take $\alpha_0 = 1$. Let $\Sigma(n) = \inf\{|\sum_{l=1}^{L}c_l\alpha_l - (n - m)| : c_l \in \mathbb{Z}^+, \sum_{l=1}^{L}c_l = m \leq n\}$. Then $\mu_n^\tau f(x) \geq \frac{1}{n}\Sigma(n)$. Thus, $O_n^\tau f \geq \frac{1}{n}\Sigma(n)$. For each such choice of $\{\alpha_l : l = 1,\dots,L\}$, there is a constant D such that $\Sigma(n) \geq \frac{1}{n^D}$. Hence, for sufficiently large n, $O_n^\tau f \geq \frac{1}{n^{D+1}} \geq \epsilon(n)$.

Here is an argument for the lower bound on $\Sigma(n)$ above. For a background on this argument, see Marcus [9], Chapter 2, and for a similar argument see Caragiu, Zaharescu, and Zaki [5]. Fix a number field K which contains all the α_j. Then each of those linear combinations $s(n) = \sum_{l=0}^{L}c_j\alpha_j$ will be a non zero element of K which is an algebraic integer. Then the norm of $s(n)$ will be a non zero rational integer number, so in absolute value it is larger than 1. That is, $|\mathrm{Norm}_{K/\mathbb{Q}}(s(n))| \geq 1$. If K has degree D over the rational numbers, then K has D embeddings $\sigma_1,\dots,\sigma_D$ in the field of complex numbers, and we also have

$$\prod_{j=1}^{D}|\sigma_j(s(n))| = |\mathrm{Norm}_{K/\mathbb{Q}}(s(n))| \geq 1.$$

On the other hand, for each $1 \leq j \leq D$,

$$|\sigma_j(s(n))| = |c_0\sigma_j(\alpha_0) + \cdots + c_L\sigma_j(\alpha_L)| \leq n\max\{|\sigma_j(\alpha_0)|,\dots,|\sigma_j(\alpha_L)|\}.$$

By the above inequalities it follows that for each $1 \leq i \leq D$,

$$|\sigma_i(s(n))| \geq \left(\prod_{\substack{1\leq j\leq D\\ j\neq i}}|\sigma_j(s(n))|\right)^{-1} \geq \frac{c}{n^{D-1}},$$

where c is a positive constant depending only on the numbers $|\sigma_j(\alpha_i)|$, $1 \le j \le D$, $0 \le i \le L$. In particular, one of the conjugates $\sigma_i(s(n))$ is $s(n)$, so $|s(n)| \ge \frac{c}{n^{D-1}}$. This bound holds uniformly for all linear combinations $s(n) = \sum_{j=0}^{L} c_j \alpha_j$, so we indeed have the lower bound

$$|\Sigma(n)| \ge \frac{c}{n^{D-1}} \ge \frac{1}{n^D}$$

for sufficiently large n.

Combining Proposition 7.4 and Proposition 7.6, we get the following.

Proposition 7.8. *For discrete spectrum dynamical systems, if* $\epsilon_1(n) = o(1/n)$ *and* $\epsilon_2(n) > 0$ *for all* n*, there is a residual set* $\mathcal{R}$ *of mean-zero functions* $f \in L_2(X)$ *such that for all* $f \in \mathcal{R}$*, we have both* $O_n^\tau f \ge \epsilon_1(n)$ *infinitely often, and* $O_n^\tau f \le \epsilon_2(n)$ *infinitely often.*

Remark 7.9.
(a) This result shows, for example, that for discrete spectrum maps, there is a residual class of functions f such that we have $O_n^\tau f \ge \frac{1}{n \log n}$ infinitely often and $O_n^\tau f \le \frac{1}{n^n}$ infinitely often too.
(b) We do not know if Proposition 7.6 holds for all ergodic dynamical systems because we do not have estimates generally showing that $O_n^\tau f$ is much smaller than $\frac{1}{n}$ infinitely often, for say a dense set of functions f. So it is possible that there is an ergodic dynamical system such that for the generic function f, there is N such that for $n \ge N$, $O_n^\tau f \ge \frac{1}{1000n}\|f\|_2$. This seems unlikely though because such an inequality cannot hold for all f. See Remark 6.7 and Proposition 7.13.

We can also get oscillation results for $\|A_n^\tau f\|_2$ in generic cases using some of the ideas here. First, we have this well-known fact.

Proposition 7.10. *Let* $\lim_{n\to\infty} \epsilon(n) = 0$. *Suppose* τ *is ergodic. Then there is a residual set* $\mathcal{R}$ *in the mean-zero functions in* $L_2(X)$ *such that for every* $f \in \mathcal{R}$*, we have* $\|A_n f\|_2 \ge \epsilon(n)$ *infinitely often.*

Proof. For $N \ge 1$, consider $\mathcal{T}_N = \bigcap_{n \ge N} \{f \in L_2(X) : \int f\, dp = 0, \|A_n^\tau f\|_2 \le \epsilon_n\}$. It is clear that this is an L_2-norm closed set because each set $\{f \in L_2(X) : \int f\, dp = 0, \|A_n^\tau f\|_2 \le \epsilon_n\}$ is L_2-norm closed. But also $\mathcal{T}_N$ has no interior. Indeed, if it did then there would be a mean-zero $f \in L_2(X)$ and some $\delta > 0$, such that for all mean-zero $g \in L_2(X)$, with $\|g\|_2 \le 1$, we would have $f + \delta g \in \mathcal{T}_N$. But then, for all $n \ge N$ and mean-zero g, $\|g\|_2 \le 1$, we would have $\|A_n^\tau(\delta g)\|_2 \le \|A_n^\tau(f + \delta g)\|_2 + \|A_n^\tau f\|_2 \le 2\epsilon(n)$. Hence, $\|A_n^\tau g\|_2 \le 2\epsilon(n)/\delta$ for all $n \ge N$. But then, for sufficiently large n, we would have the operator norm of A_n^τ on the mean-zero functions in $L_2(X)$ being less than one, which is impossible because τ is ergodic. It follows that $\bigcup_{N \ge 1} \mathcal{T}_N$ is a set of first category. Its complement is a residual set with the property we wanted. □

But in a generic sense at least we cannot underestimate the rate $\|A_n^\tau f\|_2$ either (and so a fortiori for $O_n^\tau f$).

Proposition 7.11. *Let $\epsilon(n) > 0$ for all n. Then there is a residual set $\mathcal{R}$ in the mean-zero functions in $L_2(X)$ such that for every $f \in \mathcal{R}$, there is a residual set $\mathcal{S}$ of $\tau \in \mathcal{T}$ such that infinitely often $\|A_n^\tau f\|_2 \leq \epsilon(n)$.*

Proof. Fix an ergodic discrete spectrum mapping τ_0 such that $L_2(X)$ is spanned by eigenfunctions whose eigenvalues are roots of unity. That is, take a monothetic compact metric group G whose dual group Γ consists of roots of unity. Then let σ be group translation by a generator of G. Up to measure-theoretic isomorphism, this gives us τ_0. Now, if f is mean-zero with $f \circ \tau_0 = \gamma f$, where $\gamma^n = 1$, then $A_n^{\tau_0} f = 0$. Hence, more generally if f is mean-zero and is a finite linear combination of eigenfunctions, then infinitely often we would have $A_n^{\tau_0} f = 0$.

Now, consider the functions $\bigcup_{N\geq 1}\bigcap_{n\geq N}\{f : \|A_n^{\tau_0} f\|_2 \geq \epsilon(n)\}$. The set $\mathcal{T}_N = \bigcap_{n\geq N}\{f : \|A_n^{\tau_0} f\|_2 \geq \epsilon(n)\}$ is L_2-norm closed. But also it cannot have interior. If it did, then we could find f which is mean-zero and is a finite linear combination of eigenfunctions that is in $\mathcal{T}_N$, because these functions are dense in the mean-zero functions in $L_2(X)$ in the L_2-norm topology. But for such f, we would have for infinitely many n, $\|A_n^{\tau_0} f\|_2 = 0$ and so for some large values of n, we must have $\epsilon(n) = 0$, which it not the case. Hence, $\bigcup_{N\geq 1}\mathcal{T}_N$ is a set of first category. Let $\mathcal{R}$ be its complement.

Take a countable dense set $\mathcal{D}$ in the transformations $\mathcal{T}$. Consider the residual set $\mathcal{R}_0 = \bigcap_{\sigma\in\mathcal{D}}\mathcal{R}\sigma^{-1}$. Fix any $f \in \mathcal{R}_0$. Corresponding to this f, consider the transformations $\mathcal{S}(f) = \bigcup_{N\geq 1}\bigcap_{n\geq N}\{\tau \in \mathcal{T} : \|A_n^\tau f\|_2 \geq \epsilon(n)\}$. It is easy to see that $\mathcal{S}(f)_N = \bigcap_{n\geq N}\{\tau \in \mathcal{T} : \|A_n^\tau f\|_2 \geq \epsilon(n)\}$ is closed in the weak topology on $\mathcal{T}$. We claim it has no interior. If it did, then for some $\sigma \in \mathcal{D}$, we would have $\sigma \circ \tau_0 \circ \sigma^{-1} \in \mathcal{S}(f)_N$. But $f = F \circ \sigma^{-1}$ for some $F \in \mathcal{R}$. Using the cancelations of the conjugations, and the fact that σ is measure-preserving, we would have for $n \geq N$, $\epsilon(n) \leq \|A_n^{\sigma\circ\tau_0\circ\sigma^{-1}} f\|_2 = \|A_n^{\tau_0}(f \circ \sigma)\|_2 = \|A_n^{\tau_0} F\|_2$. But this is impossible because $\|A_n^{\tau_0} F\|_2 < \epsilon(n)$ for infinitely many n. Hence, $\mathcal{S}(f)$ is a set of first category. This shows that for all $f \in \mathcal{R}_0$, we have a residual set $\mathcal{S}(f)^c \subset \mathcal{T}$, such that for all $\tau \in \mathcal{S}(f)^c$, we have $\|A_n^\tau f\|_2 < \epsilon(n)$ infinitely often. □

We can use these results to address partially the question of oscillation for $\|A_n^\tau f\|_2$.

Proposition 7.12. *Take two sequences $\epsilon_i(n), i = 1, 2$ with $\epsilon_2(n) > 0$ for all n, and with $\lim_{n\to\infty}\epsilon_1(n) = 0$. Then for a residual set $\mathcal{S}$ of maps $\tau \in \mathcal{T}$, there is a residual set $\mathcal{R}(\tau)$ of mean-zero functions such that for infinitely many n we have $\|A_n^\tau f\|_2 \geq \epsilon_1(n)$, and for infinitely many n we have $\|A_n^\tau f\|_2 \leq \epsilon_2(n)$.*

Proof. Proposition 7.11 gives a residual set $\mathcal{R}_0$ of mean-zero functions such that for each $f \in \mathcal{R}_0$, there is some residual set $\mathcal{S}(f)$ of maps such that infinitely many n we have $\|A_n^\tau f\|_2 \le \epsilon_2(n)$. Using the Kuratowski–Ulam Theorem again, this shows that there is a residual set $\mathcal{S}$ of ergodic maps such that for each $\tau \in \mathcal{S}$, there is a residual set $\mathcal{R}_2(\tau)$ of mean-zero functions such that for infinitely many n we have $\|A_n^\tau f\|_2 \le \epsilon_2(n)$. But also, by Proposition 7.10, for each $\tau \in \mathcal{S}$, there is a residual set of mean-zero functions $\mathcal{R}_1(\tau)$ such that for each $f \in \mathcal{R}_1(\tau)$, for infinitely many n, we have $\|A_n^\tau f\|_2 \ge \epsilon_1(n)$. Let $\mathcal{R}(\tau) = \mathcal{R}_1(\tau) \cap \mathcal{R}_2(\tau)$ for each $\tau \in \mathcal{S}$. This gives the result. □

Since $O_n^\tau f \le \|A_n^\tau f\|_2$, Proposition 7.11 tells us that the generic τ has $O_n^\tau f$ small infinitely often for a generic class of functions. Combining this with Proposition 7.4 gives a result like Proposition 7.12 for $O_n^\tau f$ if we take the value of $\epsilon_1(n) = o(1/n)$. Actually, Proposition 7.6 could be used in the style of the proof of Proposition 7.12 to get the lower bound aspect of this too.

Proposition 7.13. *Take two sequences $\epsilon_j(n)(i) > 0$ for $j = 1, 2$, with $\lim_{n\to\infty} \epsilon(n)(i) = 0$ for $i = 1, 2$. Assume $\epsilon_1(n) = o(1/n)$. Then for a residual set of maps $\tau \in \mathcal{T}$, there is a residual set $\mathcal{R}(\tau)$ of mean-zero functions such that for infinitely many n we have $O_n^\tau f \ge \epsilon_1(n)$, and for infinitely many n we have $O_n^\tau f \le \epsilon_2(n)$.*

7.2 Comparison of $O_n^\tau f$ and $\|A_n^\tau f\|_2$

Now we return to the question, “How close can we get $O_n^\tau f$ and $\|A_n^\tau f\|_2$, and when?” We will again use the dual method provided by Kuratowski–Ulam Theorem. We start by considering an ergodic transformation τ_0, which is discrete spectrum and has all eigenvalues being roots of unity that are a power of 2. This transformation has the property that there is a set $\mathcal{D}$ of mean-zero functions, dense in L_2-norm in the mean-zero functions in $L_2(X)$, such that these functions have the form $f = \sum_{l=1}^m d_l \chi_l$, where m is a power of 2, $d_l \in \mathbb{C}$, and $\chi_l \circ \tau_0 = \exp(2\pi i \frac{l}{m})\chi_l$ for all l. For these functions, we have $A_m^{\tau_0} f = 0$ because for each $l = 1, \dots, m$, we have $A_m^{\tau_0} \chi_l = 0$. Indeed, for each l, $\exp(2\pi i \frac{l}{m})$ is an m-th root of unity, and so $\sum_{k=1}^m \exp(2\pi i \frac{lk}{m}) = 0$. Hence, $A_{m+1}^{\tau_0} f = \frac{1}{m+1} f \circ \tau_0^{m+1}$ and so $\|A_{m+1}^{\tau_0} f\|_2 = \frac{1}{m+1}\|f\|_2$. We claim also the following.

Proposition 7.14. *For $f \in \mathcal{D}$, $f = \sum_{l=1}^m d_l \chi_l$, we have $O_{m+1}^{\tau_0} f = \frac{1}{m+1}\|f\|_2$.*

Proof. We know that $\|A_{m+1}^{\tau_0} f\|_2 = \frac{1}{m+1}\|f\|_2$ because $A_m^{\tau_0} f = 0$. So we need to only show that $O_{m+1}^{\tau_0} f \ge \frac{1}{m+1}\|f\|_2$. Take a representative measure $\mu_{m+1} = \frac{1}{m+1}\sum_{k=1}^{m+1} \delta_{m_k}$. Then $\mu_{m+1}^{\tau_0} f = \sum_{l=1}^m d_l \mu_{m+1}^{\tau_0}(\chi_l) = \sum_{l=1}^m d_l (\frac{1}{m+1}\sum_{k=1}^{m+1} \exp(2\pi i l m_k/m))\chi_l$. So, orthogonality of χ_l implies that it is enough to show that

$|\sum_{k=1}^{m+1} \exp(2\pi i l m_k/m)| \geq 1$ for all l. The sum $\sum_{k=1}^{m+1} \exp(2\pi i l m_k/m) = \sum_{j=1}^{m} c_j \exp(2\pi i j/m)$, where c_j are positive integers with $\sum_{j=1}^{m} c_j = m+1$. Consider the vectors in $\mathbb{C}^m$, $v_1 = (1,\ldots,1)$, $v_2 = (c_j : j = 1,\ldots,m)$, and $v_3 = (\exp(2\pi i(l/m) : l = 1,\ldots,m)$. We have v_1 perpendicular to v_3 and $\langle v, v_1\rangle = m+1 = \langle v_1, v_2\rangle$. So to minimize $|\langle v_2, v_3\rangle|$, we need to take v_2 as close to the $\mathcal{L}$, the line through the origin in the direction of v_1, as possible. That is, we need to minimize $\sum_{j=1}^{m} |c_j - 1|^2 = \sum_{j=1}^{m} c_j^2 - 2(m+1) + m$. It is easy to see that this occurs when all $c_j = 1$, except one of them which is 2. This gives $|\sum_{j=1}^{m} c_j \exp(2\pi i j/m)| \geq 1$ because $\sum_{j=1}^{m} \exp(2\pi i j/m) = 0$. □

Using the same basic category argument as we have been using with appropriate modifications gives the following result.

Proposition 7.15. *Given any $\epsilon(n) > 0$, there is a residual set $\mathcal{R}$ of mean-zero functions such that for each $f \in \mathcal{R}$, there are infinitely many n such that $O_n^{\tau_0} f > (1-\epsilon(n))\|A_n^{\tau_0} f\|_2$.*

This in turn proves this approximation result.

Proposition 7.16. *Given $\epsilon(n) > 0$, there is a residual set of maps $\mathcal{S} \subset \mathcal{T}$ such that for each $\tau \in \mathcal{S}$, there is a residual set of mean-zero functions $\mathcal{R}(\tau)$ such that $f \in \mathcal{R}(\tau)$, for any $\epsilon > 0$, there are infinitely many n such that $O_n^{\tau} f > (1-\epsilon(n))\|A_n^{\tau} f\|_2$.*

Proof. First, proceeding as in Proposition 7.11, we take a countable dense set $\mathcal{D}$ in the transformations $\mathcal{T}$. We take $\mathcal{R}$ from Proposition 7.15. Then consider the residual set $\mathcal{R}_0 = \bigcap_{\sigma\in\mathcal{D}} \mathcal{R}\sigma^{-1}$. Fix any $f \in \mathcal{R}_0$. Corresponding to this f, consider the transformations $\mathcal{S}(f) = \bigcup_{N\geq 1}\bigcap_{n\geq N}\{\tau \in \mathcal{T} : O_n^{\tau} f \leq (1-\epsilon(n))\|A_n^{\tau} f\|_2\}$. We will show that this is a set of first category and so its complement is residual.

But it is easy to see that $\mathcal{S}(f)_N = \bigcap_{n\geq N}\{\tau \in \mathcal{T} : O_n^{\tau} f \leq (1-\epsilon(n))\|A_n^{\tau} f\|_2\}$ is closed in the weak topology on $\mathcal{T}$. We claim it has no interior. If it did, then for some $\sigma \in \mathcal{D}$, we would have $\sigma\circ\tau_0\circ\sigma^{-1} \in \mathcal{S}(f)_N$. Now $f = F\circ\sigma^{-1}$ for some $F \in \mathcal{R}$. Using the cancelations of the conjugations, and the fact that σ is measure-preserving, we would have $\|A_n^{\sigma\circ\tau_0\circ\sigma^{-1}} f\|_2 = \|A_n^{\tau_0}(f\circ\sigma)\|_2 = \|A_n^{\tau_0} F\|_2$. Similarly, $O_n^{\sigma\circ\tau_0\circ\sigma^{-1}} f = O_n^{\tau_0} F$. This cannot happen for all $n \geq N$ because $O_n^{\tau_0} F > (1-\epsilon(n))\|A_n^{\tau_0} F\|_2$ for infinitely many n. Hence, $\mathcal{S}(f)$ is a set of first category.

As above, we can use the Kuratowski–Ulam Theorem to reverse the roles of f and τ. This gives this generic result. □

Remark 7.17. It is interesting that discrete spectrum mappings with eigenvalues all roots of unity play two very different roles in this article. This class is the one unsolved case for answering whether generically the optimal norm $O_n^{\tau} f$ is much less than $\|A_n^{\tau} f\|_2$. Essentially, this is because of the features noted in Proposition 7.14, and before it. But, on the other hand, it is these very properties that give the fact that the generic transformation has $O_n^{\tau} f$ very close to $\|A_n^{\tau} f\|_2$ infinitely often, for the generic function. This dual role for this class of transformations may be an accident of our arguments rather than an actual feature of dynamical systems.

Using the type of argument in Proposition 7.16 and Proposition 4.7, one can prove that the generic transformation satisfies the property of Proposition 4.7. Combining this with Proposition 7.16 itself gives the following simultaneous statement of oscillation.

Proposition 7.18. *Given K_n and $\epsilon(n) > 0$, there is a residual set of transformations $\mathcal{S}$ such that for each $\tau \in \mathcal{S}$, there is a residual set of mean-zero functions $\mathcal{R}(\tau)$ such that $f \in \mathcal{R}(\tau)$, there are infinitely many n such that $O_n^\tau f > (1 - \epsilon(n))\|A_n^\tau f\|_2$, and there are infinitely many n such that $K_n O_n^\tau f < \|A_n^\tau f\|_2$.*

Remark 7.19. In all of the results above that are generic results, the question can be asked if the same result is not true for all the functions, or transformations as the case may be. In situations where the answer to this is negative, then of course the issue is then whether there is a dense class of exceptions. We have tried to answer this type of question wherever possible, but we do not have at this time a complete picture in all cases. For example, in Proposition 7.18, it may be that the result is true not just for the generic transformation, but for all transformations.

8 Cocycles

Some of the result in Section 7 can be used to make some interesting observations about the norms of cocycles. First, there are a few background results that make clear what the issues are. Suppose that τ is ergodic and $f \in L_2(X)$. Consider the cocycle $S_n^\tau f = \sum_{k=1}^n f \circ \tau^k$. The following result is well-known. It appeared first for power-bounded operators in reflexive spaces in Browder [2]. This has been extended too, e.g., see Lin and Sine [8]. A proof of the following for general abelian groups appears in Parry and Schmidt [11]. We give a proof of this in our basic context for completeness, and for reference in the sequel.

Proposition 8.1. *We have* $\sup_{n\geq 1} \|S_n^\tau f\|_2 < \infty$ *if and only if f is a coboundary, i.e., there exists $g \in L_2(X)$ such that $f = g - g \circ \tau$.*

Proof. Since $n|\int f\, dp| \leq |\int S_n^\tau f\, dp| \leq \|S_n^\tau f\|_2$, we see that $\sup_{n\geq 1} \|S_n^\tau f\|_2 < \infty$ implies that $\int f\, dp = 0$. Also, clearly $S_n^\tau f - S_n^\tau f \circ \tau = f \circ \tau - f \circ \tau^{n+1}$. So $\frac{1}{N}\sum_{n=1}^N (S_n^\tau f - S_n^\tau f \circ \tau) = f \circ \tau - \frac{1}{N}\sum_{n=1}^N f \circ \tau^{n+1}$. The Mean Ergodic Theorem says that $\|\frac{1}{N}\sum_{n=1}^N f \circ \tau^{n+1}\|_2 \to 0$ as $N \to \infty$. But at the same time, $\sup_{n\geq 1} \|S_n^\tau f\|_2 < \infty$ implies that along some subsequence of values N_m, $\frac{1}{N_m}\sum_{k=1}^{N_m} S_n^\tau f$ converges weakly in $L_2(X)$ to some function $h \in L_2(X)$. But then, $\frac{1}{N_m}\sum_{n=1}^{N_m} S_n^\tau f \circ \tau$ converges weakly to $h \circ \tau$. Hence, $h - h \circ \tau = f \circ \tau$, and so $f = (h + f) - (f + h) \circ \tau$. □

Remark 8.2. The same result holds for power-bounded operators, by essentially the same proof.

We also can show the following in the same style.

Proposition 8.3. *Suppose that τ is weakly mixing and* $\sup_{n\in\mathcal{N}} \|S_n^\tau f\|_2 < \infty$ *for some set $\mathcal{N}$ that has positive lower density. Then f is a coboundary.*

Proof. Using values of $n \in \mathcal{N}$, one shows as above that f is mean-zero. Then because τ is weakly mixing, for any $g \in L_2(X)$, we know that $\frac{1}{N}\sum_{n=1}^{N} |\langle g, f \circ \tau^n\rangle| \to 0$ as $N \to \infty$. Hence, $\langle g, f \circ \tau^n\rangle$ converges to 0 along values n in a sequence $\mathcal{Z}_g$ of density one. Because our probability space is a standard Lebesgue space, a short argument shows that indeed $f \circ \tau^n$ converges to 0 weakly in $L_2(X)$ along values n in a sequence $\mathcal{Z}$ of density one. Let $\mathcal{N}_0 = \mathcal{Z} \cap \mathcal{N}$. Then $\mathcal{N}_0$ has positive lower density, indeed the same lower density as $\mathcal{N}$. Now restrict ourselves to values $n \in \mathcal{N}_0$, and use $S_n^\tau f - S_n^\tau f \circ \tau = f \circ \tau - f \circ \tau^{n+1}$. Our boundedness assumption shows that some subsequence of the left-hand side converges weakly to $h - h \circ \tau$ for some $h \in L_2(X)$, while the right-hand side is converging weakly to $f \circ \tau$. But then, $f = (f + h) - (f + h) \circ \tau$ and so f is a coboundary. □

But even fewer terms can be in $\mathcal{N}$ if we assume more about τ. We will be using the following property:

(B) For a strictly increasing sequence of whole numbers (n_m),

$$\sup_{m\geq 1} \|S_{n_m}^\tau f\|_2 < \infty.$$

Here is the result that was essentially proved by Browder and Petryshyn [3].

Proposition 8.4. *Suppose τ is strongly mixing. Then (B) implies that f is a coboundary.*

Proof. Using the boundedness along values n_m, we can see as above that f is mean-zero. Also, (B) implies that some subsequence of $S_{n_m} f$ converges weakly in $L_2(X)$ to some g. But then $g - g \circ \tau$ is the weak limit in $L_2(X)$ of $S_{N_m} f - S_{N_m} f \circ \tau = f - f \circ \tau^{n_m+1}$. Since f is mean-zero and τ is strongly mixing, we know that $f \circ \tau^{n_m}$ converges weakly to 0 in $L_2(X)$. Hence, $g - g \circ \tau = f$ and so f is a coboundary. □

Remark 8.5. So if τ is strongly mixing and f is not a coboundary, then $\|S_n^\tau f\|_2$ goes to infinity. But how fast does $\|S_n^\tau f\|_2$ go to infinity? For example, given $a_n = o(n)$, does there exist a mean-zero f such that $\|S_n^\tau f\|_2 \geq a_n$ for all n? There are some constraints. In the extreme case, if $\|S_n^\tau f\|_2 = n\|f\|_2$ for some n, and τ is ergodic, then f is constant. But also, for any non zero, mean-zero f, there exists $\rho < 1$ such that if $\|S_n^\tau f\|_2 \leq \rho n \|f\|_2$ for all n, because otherwise one can use the uniform convexity of $L_2(X)$ and show that f is constant, which it is not.

These results suggest this basic question.

Question: For which transformations τ is it always the case that f is a coboundary if (B) holds? Alternatively, for which transformations is there a mean-zero function f for which (B) holds, but f is not a coboundary?

It is not clear if there is any ergodic transformation other than the strongly mixing transformations for which (B) always implies that the function is a coboundary. But we can say more in some generality. In particular, it follows from Proposition 8.8 that the class for which (B) always implies that the function is a coboundary must be a set of first category. First, we have this restatement of a previous observation.

Proposition 8.6. *Assume that* τ *is ergodic and* $\lim_{n\to\infty} \epsilon(n) = 0$. *Then there is a residual set* $\mathcal{R}(\tau)$ *of mean-zero functions such that for infinitely many* n *we have* $\|S_n^\tau f\|_2 \geq n\epsilon_1(n)$.

Proof. This is a direct consequence of Proposition 7.10. □

Remark 8.7. This of course says that the coboundaries, which form a dense subspace, are also a first category class of functions.

On the other hand, at least generically we can get an upper estimate too.

Proposition 8.8. *Take a sequence* $\epsilon(n) > 0$ *for all* n. *Then for a residual set* $\mathcal{S}$ *of maps* $\tau \in \mathcal{T}$, *there is a residual set* $\mathcal{R}(\tau)$ *of mean-zero functions such that for infinitely many* n *we have* $\|S_n^\tau f\|_2 \leq \epsilon(n)$.

Proof. We clearly use Proposition 7.12 to get this result. One just takes $\epsilon(n)/n$ as the value for $\epsilon_2(n)$ in Proposition 7.12. □

It is not clear what the class of transformations is in Proposition 8.8. But if we combine Proposition 8.8 and Proposition 8.6, we have the following.

Corollary 8.9. *Take two sequences* $\epsilon_i(n), i = 1, 2$ *with* $\epsilon_2(n) > 0$ *for all* n, *and with* $\lim_{n\to\infty} \epsilon_1(n) = 0$. *Then for a residual set* $\mathcal{S}$ *of maps* $\tau \in \mathcal{T}$, *there is a residual set* $\mathcal{R}(\tau)$ *of mean-zero functions such that for infinitely many* n *we have* $\|S_n^\tau f\|_2 \geq n\epsilon_1(n)$, *and for infinitely many* n *we have* $\|S_n^\tau f\|_2 \leq \epsilon_2(n)$.

Remark 8.10. Derriennic and Lin [6] have some results that are of this type. While they are not as strong as what is shown here, they do obtain their results for a general ergodic mapping τ.

We can also prove a rate result for a general discrete spectrum mapping, although it is not as good as the generic case above (or even the case where the map is discrete spectrum with all eigenvalues being roots of unity). Here is what we can show.

Proposition 8.11. *Suppose τ is ergodic and has a discrete spectrum. Then there is a residual set of mean-zero functions f in $L_2(X)$ such that for all $\epsilon > 0$, we have $\|S_n^\tau f\|_2 < \epsilon$ for infinitely many values of n.*

Proof. Consider the set

$$\mathcal{E} = \bigcup_{K\geq 1} \bigcup_{N\geq 1} \bigcap_{n\geq N} \{f \in L_2(X) : \int f\,dp = 0 \text{ and } \|S_n^\tau f\|_2 \geq 1/K\}.$$

Each $\{f \in L_2(X) : \int f\,dp = 0 \text{ and } \|S_n^\tau f\|_2 \geq 1/K\}$ is closed in L_2-norm and so then is $\mathcal{E}(N,K) = \bigcap_{n\geq N}\{f \in L_2(X) : \int f\,dp = 0 \text{ and } \|S_n^\tau f\|_2 \geq 1/K\}$. We claim this set has no interior. This proves that $\mathcal{E}$ is a set of first category. So its complement $\mathcal{E}^c$ is residual and gives the class we wanted.

Suppose $\mathcal{E}(N,K)$ contains interior. Then there is a mean-zero function $f \in \mathcal{E}(N,K)$ which is a finite linear combination $\sum_{l=1}^L c_l \chi_l$ of non trivial eigenfunctions for τ. We may suppose that the χ_l are orthogonal. We take $\chi_l \circ \tau = \alpha_l \chi_l$. Then $S_n^\tau \chi_l = (\sum_{k=0}^{n-1} \alpha_l^k)\chi_l$. Because χ_l is non trivial, $\alpha_l \neq 1$ and $\sum_{k=0}^{n-1} \alpha_l^k = (\alpha_l^n - 1)/(\alpha_l - 1)$. Hence, $\|S_n^\tau f\|_2^2 = \sum_{l=1}^L |c_l|^2 |(\alpha_l^n - 1)/(\alpha_l - 1)|^2$. But we can choose n, as large as we like, such that all the values $|\alpha_l^n - 1|$ are simultaneously arbitrarily small. Hence, we can choose n, as large as we like, such that $\|S_n^\tau f\|_2 < \frac{1}{K}$. This contradicts $f \in \mathcal{E}(N,K)$ and so $\mathcal{E}(N,K)$ cannot have interior. $\square$

Remark 8.12. For the generic τ and f, what can be said about the density and structure of the sequences of times where $\|S_n^\tau f\|_2 > \epsilon_1(n)$, and what can be said about the density and structure of the sequences of times where $\|A_n^\tau f\|_2 < \epsilon_2(n)$?

This category theorem gives the following.

Proposition 8.13. *Given τ, which is ergodic and has discrete spectrum, there is a residual class of mean-zero functions f such that f is not a coboundary and (B) holds.*

Remark 8.14. In both Proposition 8.13 and Proposition 8.8, it is important to remember the saturation principle that we noted in Proposition 7.1. This shows that for a non zero, mean-zero $f \in L_2(X)$, we cannot have $\|S_n^\tau f\|_2 \to 0$ as $n \to \infty$. That is, for any non zero, mean-zero f, there exists $\delta > 0$ such that $\|S_n^\tau f\|_2 \geq \delta$ infinitely often. Hence, although we may not know how large $\|S_n^\tau f\|_2$ can be, and we have seen that it is infinitely often very small, still it cannot be too small for all n.

At the same time that we have the generic behavior for norms of cocycles above, we have an intrinsic behavior too.

Proposition 8.15. *Suppose τ is ergodic. Take (a_n) such that $\sum_{n=1}^\infty a_n^2/n^2 < \infty$. Then there exists a mean-zero $f_0 \in L_2(X)$ such that $\|S_n^\tau f_0\| \geq a_n$ for all n.*

Corollary 8.16. *Suppose τ is ergodic. Then there is a dense set of mean-zero functions $f \in L_2(X)$ such that for $\|S_n^\tau f\|_2 \to \infty$ as $n \to \infty$.*

Proof. Take the function f_0 constructed in Proposition 8.15 with $a_n \to \infty$. Consider any function $f = g - g \circ \tau + f_0$. Then $\|S_n^\tau f\|_2 \geq \|S_n^\tau f_0\|_2 - 2\|g\|_2$, which goes to ∞ as $n \to \infty$. Because the coboundaries span a dense subspace of the mean-zero functions, we have functions f of this form being dense in the mean-zero functions in $L_2(X)$. □

Proof of Proposition 8.15. Choose a mean-zero function $f^* \in L_2(X)$ which is of maximal spectral type among the spectral measures ν_f^τ where $f \in L_2(X)$ is mean-zero. Because τ is ergodic, $\nu_{f^*}^\tau(A) > 0$ for every non trivial arc in $\mathbb{T}$. Now for each $n \geq 1$, there is a small enough arc A_n around $\exp(2\pi i/2n)$ such that $|\gamma^n - 1| \geq 1$ and $|\gamma - 1| \leq \frac{1}{n}$ for all $\gamma \in A_n$. Now take $\beta_n > 0$ to be selected below and let the mean-zero function $f_0 \in L_2(X)$ be given by $\nu_{f_0}^\tau = \sum_{n=1}^\infty \beta_n 1_{A_n} \nu_{f^*}^\tau$. The function f_0 exists because the measure $\sum_{n=1}^\infty \beta_n 1_{A_n} \nu_{f^*}^\tau$ is absolutely continuous with respect to $\nu_{f^*}^\tau$. Also, $\|f_0\|_2^2 = \sum_{n=1}^\infty \beta_n \nu_{f^*}^\tau(A_n)$.

Now we underestimate

$$\begin{aligned}
\|S_n^\tau f_0\|_2^2 &= \int_{\mathbb{T}} \left|\frac{\gamma^n - 1}{\gamma - 1}\right|^2 d\nu_{f_0}^\tau(\gamma) \\
&\geq \int_{A_n} \beta_n \left|\frac{\gamma^n - 1}{\gamma - 1}\right|^2 d\nu_{f^*}^\tau(\gamma) \\
&\geq \int_{A_n} n^2 \beta_n d\nu_{f^*}^\tau(\gamma) \\
&= n^2 \beta_n \nu_{f^*}^\tau(A_n).
\end{aligned}$$

So choose each β_n so that $\beta_n \nu_{f^*}^\tau(A_n) = a_n^2/n^2$. Hence, $\|S_n^\tau f_0\|_2 \geq a_n$ for all n. But also $\|f_0\|_2^2 = \sum_{n=1}^\infty a_n^2/n^2 < \infty$. □

Remark 8.17. Given τ ergodic (perhaps even strongly mixing), and f not a coboundary, what rates can $\|S_n^\tau f\|_2$ go to infinity? For example, given $a_n = o(n)$, does there exist a mean-zero f such that $\|S_n^\tau f\|_2 \geq a_n$ for all n? Proposition 8.15 gives a result of this type once $\sum_{n=1}^\infty a_n^2/n^2$ converges. Simple examples of such sequences include sequences (a_n) like $\sqrt{n}/\log^{1/2+\epsilon}$ for some $\epsilon > 0$.

Also, using Derriennic and Lin [6], results like this can be proved for more general operators T in place of the Koopman operators, and also proved in other Lebesgue spaces. But it is not clear where the best underestimate for rates one can get on a dense set of functions.

9 Other Lebesgue spaces

We can derive most of the results from the previous sections in the other L_r-spaces. But there are some cases where it is not yet clear how to obtain analogous results. We define $O^\tau_{n,r}f$ to be the infimum of all $\|\mu^\tau_n f\|_r$ where $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{m_k}$ for some n choices of $m_k \in \mathbb{Z}$. We want to see how this optimal value compares to $\|A^\tau_n f\|_r$, and how both of these quantities vary as n varies for particular functions and transformations.

9.1 Optimum rarely vanishes

The first result is that typically $O^\tau_{n,r}f$ is not zero. This depends on our getting a lower estimate for $\|\mu^\tau_n f\|_r$ for suitable L_r-norm one functions.

Proposition 9.1. *For fixed n and* $1 < r < \infty$, *if* $np(E) < 1$, *we have*

$$\|\mu^\tau_n(1_E - p(E))\|_r \geq \frac{p(E)^{1/r}(1 - p(E)^{(r-1)/r} n^{(r-1)/r})}{n^{(r-1)/r}}.$$

For fixed n and $1 < r < \infty$, *if* $np(E) < 1$, *then*

$$\|\mu^\tau_n(1_E - p(E))\|_1 \geq \frac{p(E)(1 - p(E)^{(r-1)/r} n^{(r-1)/r})^r}{n^{r-1}}.$$

Proof. Notice that we have $\|\mu^\tau_n(1_E - p(E))\|_r \geq \|\mu^\tau_n 1_E\|_r - p(E)$. But

$$\begin{aligned}\|\mu^\tau_n 1_E\|_r^r &= \frac{1}{n^r}\int\left(\sum_{k=1}^n 1_{\tau^{-m_k}E}\right)^r dp\\ &\geq \frac{1}{n^r}\int\sum_{k=1}^n 1_{\tau^{-m_k}E}\,dp\\ &= \frac{1}{n^r}np(E).\end{aligned}$$

Hence, $\|\mu^\tau_n(1_E - p(E))\|_r \geq \frac{p(E)^{1/r}}{n^{(r-1)/r}} - p(E)$. This gives the lower estimate on $L_r(X)$. It follows also that

$$\frac{p(E)(1 - p(E)^{(r-1)/r} n^{(r-1)/r})^r}{n^{r-1}} \leq \int \mu^\tau_n(1_E - p(E))^r\,dp \leq \int |\mu^\tau_n(1_E - p(E))|\,dp$$

because $\|\mu^\tau_n(1_E - p(E))\|_\infty \leq 1$. This gives the underestimate for $L_1(X)$. □

Corollary 9.2. *For fixed n and* $\delta > 0$, *with* $p(E)$ *sufficiently small, and the function* $f = (1_E - p(E)/\|1_E - p(E)\|_1$, *we have*

$$O^\tau_{n,1}f \geq \frac{1}{n^\delta}.$$

Also, for any fixed $r, 1 < r < \infty$ and n, with $p(E)$ sufficiently small, and the function $f = (1_E - p(E)/\|1_E - p(E)\|_r$, we have

$$O^\tau_{n,r} f \geq \frac{1}{2n^{(r-1)/r}}.$$

Proof. We have $\|1_E - p(E)\|_1 \sim 2p(E)$ as $p(E) \to 0$, and for $1 < r < \infty$, we have $\|1_E - p(E)\|_r \sim p(E)^{1/r}$ as $p(E) \to 0$. But then the asymptotic lower estimate easily follows from this and Proposition 9.1. □

Now our averages μ^τ_n are L_r-contractions for all $r, 1 \leq r \leq \infty$. This is all that is needed together with Corollary 9.2 to prove the following analogue of Proposition 2.7.

Proposition 9.3. *Let $\mathcal{Z}$ consist of all mean-zero functions $f \in L_r(X)$ such that $O^\tau_{n,r} f = 0$ for some $n \geq 2$. Then $\mathcal{Z}$ is a set of first category.*

9.2 Optimum often smaller

We would like to be able to show that $O^\tau_{n,r} f$ is infinitely often much smaller than $\|A^\tau_n f\|_r$ for the generic function in $f \in L_r(X)$. For example, assume τ is weakly mixing. We cannot use the correlation approach that worked with $r = 2$. However, there are estimates in the IID case. For example, if $1 < r \leq 2$, then $\| \sum_{k=1}^n f_k \|_r \leq 2n^{1/r} \|f\|_r$. See Rosenthal[13], Lemma 2(b). For weakly mixing τ, we can approximate this type of estimate. We need this lemma that is perhaps of independent interest.

Lemma 9.4. *Suppose τ is weakly mixing and $1 \leq r < \infty$. Consider a non zero, mean-zero function $g \in L_r(X)$. For any $\delta > 0$, there are choices of $m_1, \ldots, m_n$ and functions $G_1, \ldots, G_n$ that are IID such that $\|G_k - g \circ \tau^{m_k}\|_r < \delta$ for all $k = 1, \ldots, n$. It can also be arranged that for all k, G_k is mean-zero and $\|G_k\|_r = \|g\|_r$.*

Proof. Suppose A, D are measurable sets with $0 < p(A) < 1$ and $0 < p(D) < 1$. Let $\gamma > 0$. Then there exists $\epsilon > 0$ such that if $|p(A \cap D) - p(A)p(D)| < \epsilon$, there is E with $p(E) = p(D)$, $p(E \Delta D) < \gamma$, and $p(A \cap E) = p(A)p(E)$. That is, if D is almost independent of A, then there is a set E close to D in measure which is independent of A. This is proved by adjusting D, i.e., adding, or taking away, sets of the same small measure. For example, if $\epsilon > 0$ is sufficiently small and $p(A)p(D) - \epsilon < p(A \cap D) < p(A)p(D)$, then one could select sets $C_1 \subset D \backslash A$ and $C_2 \subset A \backslash D$ of the same small measure so that $E = (D \backslash C_1) \cup C_2$ is independent of A, $p(E) = p(D)$, and $p(E \Delta D) < \gamma$. No matter what γ is, if ϵ is small enough, this will be possible.

But then for any $\gamma > 0$ and sets A, B with $0 < p(A) < 1$ and $0 < p(B) < 1$, since τ is weakly mixing, there exists n such that $|p(A \cap \tau^n B) - p(A)p(B)| < \epsilon$ with ϵ as above. Taking $D = \tau^n B$, there is a set E which is independent of A with $p(E) = p(B)$ and $p(\tau^m B \Delta E) < \gamma$. Indeed, in the same fashion, it is not hard to see that given $A_1, \dots, A_n, B$, with measures in $(0,1)$, and $\gamma > 0$, there is $\epsilon > 0$ sufficiently small, so that once we choose m with $|p(A_i \cap \tau^m B) - p(A_i)p(B)| < \epsilon$ for all i, there is a set E which is independent of all the A_i, but also has $p(E) = p(B)$ and $p(\tau^m B \Delta E) < \gamma$. This is accomplished by removing and adding small sets from $\tau^m B$ to achieve independence from all A_i simultaneously.

It will be important in the construction to make one additional refinement above. Suppose one is given $A_1, \dots, A_n$, and pairwise disjoint $B_1, \dots, B_N$, with measures in $(0,1)$, and some $\gamma > 0$. Then there is a sufficiently small ϵ so that once $|p(A_i \cap \tau^m B_j) - p(A_i)p(B_j)| < \epsilon$ for all i, j, there are pairwise disjoint sets E_j which are each independent of all A_i, such that also $p(E_j) = p(B_j)$ and $p(E_j \Delta \tau^m B_j) < \gamma$ for all j.

Fix $\gamma > 0$. Now, take a simple function s such that $\|s - g\|_r \le \gamma/2$. Suppose the sets $L_1, \dots, L_M$ are the pairwise disjoint sets where s takes distinct, non zero values $c_1, \dots, c_M$ respectively. Once γ is small enough, since g is non zero and mean-zero, the sets L_i would all have measure in $(0,1)$. We can also arrange for s to be mean-zero and have $\|s\|_r = \|g\|_r$.

Since τ is weakly mixing, we can use the method above to choose m_1 so that the sets $\tau^{m_1} L_j$ are all so close to being independent of each L_i that they can be replaced by pairwise disjoint sets $E_j(1)$ which are independent of each L_i while yet both $p(E_j(1)) = p(L_j)$ and $p(\tau^{m_1} L_j \Delta E_j(1)) < \gamma$ for all j. Now repeat this construction to choose m_2 so that the sets $\tau^{m_2} L_j$ are each very close to being independent of all $E_i(1)$. This allows us to choose pairwise disjoint $E_j(2)$ with $p(E_j(2)) = p(L_j)$ and $p(\tau^{m_2} L_j \Delta E_j(2)) < \gamma$ for all j, and each $E_j(2)$ is actually independent of all $E_i(1)$. Continue this pattern inductively to choose $m_1, \dots, m_n$ so that $\tau^{m_k} L_j$ is almost independent of all $E_i(l)$ with $l < k$ too. With sufficiently close approximation this allows the construction of pairwise disjoint $E_j(k)$ which are actually independent of all $E_i(l)$ with $l < k$, yet $p(E_j(k)) = p(L_j)$, and $p(\tau^{n_k} L_j \Delta E_j(k)) < \gamma$ for all j.

Now consider the functions $G_k = \sum_{i=1}^m c_i 1_{E_k(i)}$. These have been constructed to be IID. Actually, they all have the same distribution as s. In the approximations above, once γ is sufficiently small, we would have $\|G_k - g \circ \tau^{-m_k}\|_r < \delta$ for all $k = 1, \dots, n$. □

This lemma allows us to prove the following result.

Proposition 9.5. *Suppose τ is weakly mixing and $1 < r < \infty$. Then for a generic set of functions $f \in L_r(X)$, we have for all K, $O_{n,r}^\tau f < \frac{1}{K} \|A_n^\tau f\|_r$ for infinitely many n.*

Proof. First consider a fixed mean-zero function $g \in L_r(X)$ such that $1 < r \le 2$. Fix some $\epsilon > 0$. Because τ is weakly mixing, we can use Lemma 9.4 to see that there are choices of integers $m_1, \dots, m_n$ and functions $G_1, \dots, G_n$ that are IID, such that

$\|G_k - g \circ \tau^{m_k}\|_r < \epsilon$ for all $k = 1, \dots, n$. But then we have

$$\left\| \sum_{k=1}^{n} g \circ \tau^{m_k} \right\|_r \leq n\epsilon + \left\| \sum_{k=1}^{n} G_k \right\|_r \leq n\epsilon + 2n^{1/r}\|G_1\|_r \leq n\epsilon + 2n^{1/r}(\|g\|_r + \epsilon).$$

Because ϵ is arbitrary, this tells us that $O^{\tau}_{n,r}g \leq \frac{1}{n^{1-\frac{1}{r}}}\|g\|_r$.

In case $2 < r < \infty$, we use Theorem 3 in Rosenthal [13] and an argument just like the one above to show that for some constant K_r, we have $O^{\tau}_{n,r}f \leq K_p \max(\frac{1}{n^{1-\frac{1}{r}}}\|g\|_r, \frac{1}{n^{1/2}}\|g\|_2) \leq K_r \frac{1}{n^{(r-1)/r}}\|g\|_r$.

Now, as before, consider the class $\mathcal{O}_r(\mathbb{N})$ of all mean-zero functions $f \in L_r(X)$ such that for some K and some N, we have $\|A^{\tau}_n f\|_r \leq KO^{\tau}_{n,r}f$ for all $n \geq N$. This is a countable union of closed sets because each $\mathcal{O}(N,K) = \bigcap_{n\geq N}\{f \in L_r(X) : \int f\,dp = 0, \|A^{\tau}_n f\|_r \leq KO^{\tau}_{n,r}f\}$ is closed in L_r-norm. To prove this proposition, we need to only show that $\mathcal{O}(N,K)$ contains no interior. If this were not the case, then there would exist f_0 and $\epsilon_0 < 1$ such that for all mean-zero g, $\|g\|_r \leq 1$, we have $f_0 + \epsilon_0 g$ in this set. Choose $n \geq N$.

Take first the case $1 < r \leq 2$. We would have

$$\begin{aligned}
\|A^{\tau}_n(\epsilon_0 g)\|_r &\leq \|A^{\tau}_n f_0\|_r + \|A^{\tau}_n(f_0 + \epsilon_0 g)\|_r \\
&\leq KO^{\tau}_{n,r}f_0 + KO^{\tau}_{n,r}(f_0 + \epsilon_0 g) \\
&\leq \frac{K\|f_0\|_r}{n^{1-\frac{1}{r}}} + \frac{K\|f_0 + \epsilon_0 g\|_r}{n^{1-\frac{1}{r}}} \\
&\leq \frac{2K\|f_0\|_r}{n^{1-\frac{1}{r}}} + \frac{K\epsilon_0}{n^{1-\frac{1}{r}}} \\
&\leq \epsilon_0^2
\end{aligned}$$

for some particular n which is sufficiently large. But then for this value of n, $\|A^{\tau}_n g\|_r \leq \epsilon_0$ for all mean-zero g with $\|g\|_r \leq 1$. It follows that the operator norm of A^{τ}_n on the mean-zero functions is strictly less than 1. However, it is well-known that this operator norm is always equal to one.

In the case that $2 < r < \infty$, we proceed as above and conclude that for all mean-zero $g \in L_r(X)$ with $\|f\|_r \leq 1$, we have

$$\begin{aligned}
\|A^{\tau}_n(\epsilon_0 g)\|_r &\leq \|A^{\tau}_n f_0\|_r + \|A^{\tau}_n(f_0 + \epsilon_0 g)\|_r \\
&\leq KO^{\tau}_{n,r}f_0 + KO^{\tau}_{n,r}(f_0 + \epsilon_0 g) \\
&\leq \frac{K\|f_0\|_r}{n^{1/2}} + \frac{K\|f_0 + \epsilon_0 g\|_2}{n^{1/2}} \\
&\leq \frac{2K\|f_0\|_r}{n^{1/2}} + \frac{K\epsilon_0}{n^{1/2}} \\
&\leq \epsilon_0^2
\end{aligned}$$

for some particular n which is sufficiently large. But then for this value of n, $\|A^{\tau}_n g\|_r \leq \epsilon_0$ for all mean-zero g with $\|g\|_r \leq 1$. This is impossible for the same reason as before: all the norms of $\|A^{\tau}_n\|_r$ are one on the mean-zero functions. □

Remark 9.6. What is shown in the proof above is that when τ is weakly mixing and $1 < r < \infty$, then $O^{\tau}_{n,r}f \le \frac{1}{n^{(r-1)/r}}\|f\|_r$. This overestimate is best possible if one wants a homogeneous estimate. Indeed, using Proposition 9.1, for any n and $\epsilon > 0$, we can choose E with $p(E)$ small enough so that with $f = (1_E - p(E))/\|1_E - p(E)\|_r$, we have $O^{\tau}_{n,r}f \ge (1-\epsilon)\frac{1}{n^{(r-1)/r}}$. This is the analogue for $L_r(X)$ of Remark 2.5.

Proposition 9.5 does not handle the case of $r = 1$. At this time, we do not know what is the situation for this function space when τ is just assumed to be weak mixing. However, if τ is a rotation of the circle, then we can make estimates for $O^{\tau}_{n,r}f$ for trigonometric polynomials for all r. Using the analogue of previous arguments, this allows us to at least get generic results in $L_r(X)$ for all values of $r, 1 \le r < \infty$.

Proposition 9.7. *There is a generic class of transformations $\mathcal{S}_1$ such that for all $\tau \in \mathcal{S}$, there is a generic set of mean-zero functions $f \in L_1(X)$ such that we have for all K, $O^{\tau}_{n,1}f < \frac{1}{K}\|A^{\tau}_n f\|_1$ for infinitely many n.*

Proof. Fix τ_0, an ergodic rotation of the circle. The same argument that worked in $L_2(X)$ shows that for the trigonometric polynomials $\mathcal{P}$, if $f \in \mathcal{P}$, then $O^{\tau_0}_{n,1}f = 0$. This allows one to show just as in the L_2 case that there exists a generic set of mean-zero functions $f \in L_r(X)$ such that for all K, $O^{\tau_0}_{n,1}f < \frac{1}{K}\|A^{\tau_0}_n f\|_1$ for infinitely many n. We then use the translation method to obtain a smaller class of functions $\mathcal{R}_0$, which is still generic, but for each $f \in \mathcal{R}_0$, there is a generic class of transformations τ such that for all K, $O^{\tau}_{n,1}f < \frac{1}{K}\|A^{\tau}_n f\|_1$ for infinitely many n. We apply the Kuratowski–Ulam Theorem again. This gives us the generic class of transformations that we wanted. $\square$

Proposition 9.8. *There is a generic class $\mathcal{S}$ of transformations such that for all $r, 1 \le r < \infty$, there is a generic class of functions such that for all K, $O^{\tau}_{n,r}f < \frac{1}{K}\|A^{\tau}_n f\|_r$ for infinitely many n.*

Proof. Use Proposition 9.7 to produce the generic class $\mathcal{S}_1$ for the function space $L_1(X)$. Let $\mathcal{W}$ be the weak mixing transformations and let $\mathcal{S} = \mathcal{S}_1 \cap \mathcal{W}$. Because $\mathcal{W}$ is generic, this gives a generic class of transformations that works for all the function spaces $L_r(X), 1 \le r < \infty$. $\square$

9.3 Optimum often comparable

We were able to obtain generic results that show $O^{\tau}_{n,2}f$ and $\|A^{\tau}_n f\|_2$ are infinitely often close to being equal. This result depended on the analysis of the norm behavior for a particular ergodic transformation. We can attempt to prove the same result in general, but we will need to get approximations now in the L_r-norm. This may be easier to do when $r > 1$ and the Lebesgue space is uniformly convex. We should try to use the transformation τ_0 which is an ergodic mapping all of whose eigenvalues are roots of unity. We use the same notation used in Proposition 7.14. We have a particular function f, which is actually a trigonometric polynomial, such that $A_{m+1}^{\tau_0} f = \frac{1}{m+1} f \circ \tau_0^{m+1}$, so $\|A_{m+1}^{\tau_0} f\|_r = \frac{1}{m+1}\|f\|_r$. We need to prove the analogue of Proposition 7.14; that is, to show $O_{m+1,r}^{\tau_0} f = \frac{1}{m+1}\|f\|_r$ too. It is not clear that this is true in general because of how dependent the case $r = 2$ was on orthogonality arguments.

Question: Is it possible to use this method to produce a dense class of functions in $L_r(X)$ for which infinitely often the optimal L_r-norm of averages is equal (or close to) to the L_r-norm of the usual ergodic average?

If this can be done with this mapping, or an alternative map, then the rest of the arguments can be used mutatis mutandis to get similar generic results for $L_r(X), 1 \leq r < \infty$.

9.4 Oscillation for the optimal and usual norms

We can prove analogues for both $O_{n,r}^{\tau} f$ and $\|A_n \tau f\|_r$ of the oscillation results previously obtained for $r = 2$ in Section 7. Without going into great detail, here are the two basic facts.

Generalizing Proposition 7.12, we have this result.

Proposition 9.9. *Let $1 \leq r < \infty$. Take two sequences $\epsilon_i(n), i = 1, 2$ with $\epsilon_2(n) > 0$ for all n, and with $\lim_{n\to\infty} \epsilon_1(n) = 0$. Then for a residual set $\mathcal{S}$ of maps $\tau \in \mathcal{T}$, there is a residual set $\mathcal{R}(\tau)$ of mean-zero functions such that for infinitely many n we have $\|A_n^{\tau} f\|_r \geq \epsilon_1(n)$, and for infinitely many n we have $\|A_n^{\tau} f\|_r \leq \epsilon_2(n)$.*

Proof. If we fix an ergodic τ, then it is a standard fact that on the mean-zero functions in $L_r(X)$, $\|A_n \tau\|_r = 1$. This allows one to show that for the generic mean-zero function $f \in L_r(X)$, we have $\|A_n^{\tau} f\|_r \geq \epsilon_1(n)$ infinitely often.

In addition, using the map τ_0 that we have used above, we would have $\|A_n^{\tau_0} f\|_r = 0$ infinitely often for a dense class of mean-zero functions $f \in L_r(X)$. This can be used to show that for the generic transformation τ, we have for the generic mean-zero function $f \in L_r(X)$, infinitely often $\|A_n^{\tau} f\|_r \leq \epsilon_2(n)$. □

A similar result holds for the optimal norm values with some modification on the lower estimate using Proposition 9.1.

Proposition 9.10. *Let $1 \leq r < \infty$. Take two sequences $\epsilon_i(n), i = 1, 2$. When $r = 1$, assume $\delta > 0$ and $\epsilon_1(n) = o(1/n^{2\delta})$. When $1 < r < \infty$, assume $\epsilon_1(n) = o(1/n^{2(r-1)/r})$.*

Also, assume $\epsilon_2(n) > 0$ for all n. Then for a residual set $\mathcal{S}$ of maps $\tau \in \mathcal{T}$, there is a residual set $\mathcal{R}(\tau)$ of mean-zero functions such that for infinitely many n we have $O^\tau_{n,r}f \geq \epsilon_1(n)$, and for infinitely many n we have $O^\tau_{n,r}f \leq \epsilon_2(n)$.

Proof. The estimate involving $\epsilon_2(n)$ is directly analogous to the previous arguments; indeed, this gives the corresponding estimate in Proposition 9.9. On the other hand, the estimate involving $\epsilon_1(n)$ is carried out as in the case $r = 2$, except that one uses the estimates in Proposition 9.1 in place of Proposition 2.4. □

Remark 9.11. The result in Remark 7.7 extends to L_r-norms.

Remark 9.12. In conclusion, we should observe that if we chose to use the more general optimal norm $C^\tau_{n,r}f$, the issues we have considered become substantially different. Indeed, it is easy to see that $C^\tau_{n+1,r} \leq C^\tau_{n,r}f$. Here is the effect of this when $r = 2$. Even for the case that was previously a problem, where τ has discrete spectrum and all eigenvalues are roots of unity, we can show that $C^\tau_{n,2}f$ is infinitely often much smaller than $\|A^\tau_n f\|_2$ for the generic mean-zero function in $L_2(X)$, at least in the initial basic comparison that was used above. However, it is not clear when and how these two norm gauges are comparable, without both of them being zero. This issue needs further work, as do the parallel investigation for a general value of r.

Acknowledgments

I would like to thank Bruce Reznick for giving me the nice example in Section 6, Terry Adams for the example in Remark 7.5, Alexandru Zaharescu for the argument in Remark 7.7, Michael Lin for help with Section 8, and Vitaly Bergelson for his thoughtful questions. I would also like to thank the referee for a number of good suggestions that improved this article.

References

[1] V. Bergelson and J. Rosenblatt, Mixing actions of groups, *Illinois J. Math.* **32** (1988), no. 1, 65–80.
[2] F. Browder, On iteration of transformations in non-compact minimal dynamical systems, *Proc. Amer. Math. Soc.* **9** (1958), 773–780.
[3] F. Browder and W. Petryshyn, The solution by iteration of linear functional equations in Banach spaces, *Bull. Amer. Math. Soc.* **72** (1966), 566–570.
[4] P. L. Butzer and U. Westphal, The mean ergodic theorem and saturation, *Indiana Univ. Math. J.*, **20** (1971), 1163–1174.
[5] M. Caragiu, A. Zaharescu and M. Zaki, On Ducci sequences with algebraic numbers, *Fibonacci Quart.* **49** (2011), no. 1, 34–40.

[6] Y. Derriennic and M. Lin, Fractional Poisson equations and ergodic theorems for fractional coboundaries, *Israel J. Math.* **123** (2001), 93–130.

[7] E. Hewitt and K. Ross, *Abstract Harmonic Analysis*, Vol. I, Spring-Verlag, Berlin-Göttingen-Heidelberg, 1963.

[8] M. Lin and R. Sine, Ergodic theory and the functional equation $(I - T)x = y$, *J. Operator Theory* **10** (1983), 153–166.

[9] D. Marcus, *Number Fields*, Springer-Verlag, New York-Heidelberg, 1977. viii+279 pp.

[10] J. Oxtoby, *Measure and Category*, 2nd edition, Spring-Verlag, New York-Heidelberg-Berlin, 1980.

[11] W. Parry and K. Schmidt, A note on cocycles of unitary representations, *Proc. Amer. Math. Soc.* **55** (1976), no. 1, 185–190.

[12] J. Rosenblatt, Joint coboundaries, preprint, 7 pages.

[13] H. Rosenthal, On the subspaces of $L^p (p > 2)$ spanned by sequences of independent random variables, *Israel J. Math.* **8** (1970), 273–303.

Ethan Akin

The iterated Prisoner's Dilemma: good strategies and their dynamics

Abstract: For the iterated Prisoner's Dilemma, there exist Markov strategies that solve the problem when we restrict attention to the long-term average payoff. When used by both players, these assure the cooperative payoff for each of them. Neither player can benefit by moving unilaterally any other strategy, i.e., these are Nash equilibria. In addition, if a player uses instead an alternative that decreases the opponent's payoff below the cooperative level, then his own payoff is decreased as well. Thus, if we limit attention to the long-term payoff, these *good strategies* effectively stabilize cooperative behavior. We characterize these good strategies and analyze their role in evolutionary dynamics.

Keywords: Prisoner's Dilemma, stable cooperative behavior, iterated play, Markov strategies, zero-determinant strategies, Press–Dyson equations, evolutionary game dynamics.

1 The iterated Prisoner's Dilemma

The *Prisoner's Dilemma* is a two-person game that provides a simple model of a disturbing social phenomenon. It is a symmetric game in which each of the two players, X and Y, has a choice between two strategies, c and d. Thus, there are four outcomes that we list in the order: cc, cd, dc, dd, where, for example, cd is the outcome when X plays c and Y plays d. Each then receives a payoff. The following 2×2 chart describes the payoff to the X player. The transpose is the Y payoff.

X/Y	c	d
c	R	S
d	T	P

(1.1)

Alternatively, we can define the *payoff vectors* for each player by

$$\mathbf{S}_{\mathrm{X}} = (R, S, T, P) \quad \text{and} \quad \mathbf{S}_{\mathrm{Y}} = (R, T, S, P). \tag{1.2}$$

Davis [6] and Straffin [17] provide clear introductory discussions of the elements of game theory.

Either player can use a *mixed strategy*, randomizing by choosing c with probability p_c and d with the complementary probability $1 - p_c$.

A probability *distribution* $\mathbf{v}$ on the set of outcomes is a non-negative vector with unit sum, indexed by the four states. That is, $v_i \geq 0$ for $i = 1, \ldots, 4$ and the dot product $\langle \mathbf{v} \cdot \mathbf{1} \rangle = 1$. For example, v_2 is the probability that X played c and Y played d. In particular, $v_1 + v_2$ is the probability X played c. With respect to $\mathbf{v}$, the expected payoffs to X and Y, denoted s_X and s_Y, are the dot products with the corresponding payoff vectors:

$$s_X = \langle \mathbf{v} \cdot \mathbf{S}_X \rangle \quad \text{and} \quad s_Y = \langle \mathbf{v} \cdot \mathbf{S}_Y \rangle. \tag{1.3}$$

The payoffs are assumed to satisfy

$$T > R > P > S \quad \text{and} \quad 2R > T + S. \tag{1.4}$$

We will later use the following easy consequence of these inequalities.

Proposition 1.1. *If* $\mathbf{v}$ *is a distribution, then the associated expected payoffs to the two players, as defined by Equation (1.3), satisfy the following equation:*

$$s_Y - s_X = (v_2 - v_3)(T - S). \tag{1.5}$$

So we have $s_Y = s_X$ *iff* $v_2 = v_3$.

In addition,

$$\frac{1}{2}(s_Y + s_X) \leq R, \tag{1.6}$$

with equality iff $\mathbf{v} = (1, 0, 0, 0)$. *Hence, the following statements are equivalent.*
(i) $1/2(s_Y + s_X) = R$,
(ii) $v_1 = 1$,
(iii) $s_Y = s_X = R$.

Proof. Dot $\mathbf{v}$ with $\mathbf{S}_Y - \mathbf{S}_X = (0, T - S, S - T, 0)$ and with $(1/2)(\mathbf{S}_Y + \mathbf{S}_X) = (R, (1/2)(T + S), (1/2)(T + S), P)$. Observe that R is the maximum entry of the latter. □

In the Prisoner's Dilemma, the strategy c is *cooperation*. When both players cooperate, they each receive the reward for cooperation ($=R$). The strategy d is *defection*. When both players defect, they each receive the punishment for defection ($=P$). However, if one player cooperates and the other does not, then the defector receives the large temptation payoff ($=T$), while the hapless cooperator receives the very small sucker's payoff ($=S$). The condition $2R > T + S$ says that the reward for cooperation is larger than the players would receive by dividing equally the total payoff of a *cd* or *dc* outcome. Thus, the maximum total payoff occurs uniquely at *cc* and that location is a *strict Pareto optimum*, which means that at every other outcome at least one player does worse. The cooperative outcome *cc* is clearly where the players "should" end up. If they could negotiate a binding agreement in advance of play, they would agree to play c and each receive R. However, the structure of the game is such that, at the time of play, each chooses a strategy in ignorance of the other's choice.

This is where it gets ugly. In game theory lingo, the strategy *d strictly dominates* strategy c. This means that, whatever Y's choice is, X receives a larger payoff by playing *d* than by using *c*. In Array (1.1), each number in the *d* row is larger than the corresponding number in the *c* row above it. Hence, X chooses *d*, and for exactly the same reason, Y chooses *d*. So they are driven to the *dd* outcome with payoff *P* for each. Having firmly agreed to cooperate, X hopes that Y will stick to the agreement because X can then obtain the large payoff *T* by defecting. Furthermore, if he were not to play *d*, then he risks getting *S* when Y defects. All the more reason to defect, as X realizes Y is thinking the same thing.

The payoffs are often stated in money amounts or in years reduced from a prison sentence (the original "prisoner" version), but it is important to understand that the payoffs are really in units of *utility*. That is, the ordering in Equation (1.4) is assumed to describe the order of desirability of the various outcomes to each player when all the consequences of each outcome are taken into account. Thus, if X is induced to feel guilty at the *dc* outcome, then the payoff to X of that outcome is reduced. Adjusting the payoffs is the classic way of stabilizing cooperative behavior. Suppose prisoner X walks out of prison, free after defecting, having consigned Y, who played *c*, to a 20-year sentence. Colleagues of Y might well do X some serious damage. Anticipation of such an event considerably reduces the desirability of *dc* for X, perhaps to well below *R*. If X and Y each have threatening friends, then it is reasonable for each to expect that a prior agreement to play *cc* will stand and so they each receive *R*. However, in terms of utility this is no longer a Prisoner's Dilemma. In the book that originated modern game theory, Von Neumann and Morgenstern [19], the authors developed an axiomatic theory of utility that allows us to make sense of such arithmetic relationships as the second inequality in Equation (1.4). We won't consider this further, but the reader should remember that the payoffs are numerical measurements of desirability.

This two-person collapse of cooperation can be regarded as a simple model of what Garret Hardin [7] calls *the tragedy of the commons*. This is a similar sort of collapse of mutually beneficial cooperation on a multi person scale.

In the search for a way to avert this tragedy, attention has focused upon *repeated play*. X and Y play repeated rounds of the same game. For each round, the players' choices are made independently, but each is aware of all of the previous outcomes. The hope is that the threat of future retaliation will rein in the temptation to defect in the current round.

Robert Axelrod devised a tournament in which submitted computer programs played against one another. Each program played a fixed, but unknown, number of rounds against each of the competing programs, and the resulting payoffs were summed. The results are described and analyzed in his landmark book [4]. The winning program, Tit-for-Tat, submitted by game theorist Anatol Rapaport, consists, after initial cooperation, in playing in each round the strategy used by the opponent in the previous round. A second tournament yielded the same winner. Axelrod extracted some interesting rules of thumb from Tit-for-Tat and applied these to some historical examples.

At around the same time, game theory was being introduced by John Maynard Smith into biology in order to study problems in the evolution of behavior. Maynard Smith [11] and Sigmund [13] provide good surveys of the early work. Tournament play for games, which has been widely explored since, exactly simulates the dynamics examined in this growing field of evolutionary game theory. However, the tournament/evolutionary viewpoint changes the problem in a subtle way. In evolutionary game theory, what matters is how a player is doing as compared with the competing players. Consider this with just two players and suppose they are currently considering strategies with the same payoff to each. Comparing outcomes, Y would reject a move to a strategy where she does better, but which allows X to do still better than she. That this sort of *altruism* is selected against is a major problem in the theory of evolution. However, in classical game theory, the payoffs are in utilities. Y simply desires to obtain the highest absolute payoff. The payoffs to her opponent are irrelevant, except as data to predict X's choice of strategy. It is the classical problem that we will mainly consider, although we will return to evolutionary dynamics in the last section.

I am not competent to summarize the immense literature devoted to these matters. I recommend the excellent book-length treatments of Hofbauer and Sigmund [9] and Nowak [12] and Sigmund [14]. The latter two discuss the Markov approach that we now examine.

The choice of play for the first round is the *initial play*. A *strategy* is a choice of initial play together with what we will call a *plan*: a choice of play, after the first round, to respond to any possible past history of outcomes in the previous rounds. A memory-one plan bases its response entirely on outcome of the previous round. The Tit-for-Tat plan (hereafter, just TFT) is an example of a memory-one plan.

With the outcomes listed in order as cc, cd, dc, dd, a memory one plan for X is a vector $\mathbf{p} = (p_1, p_2, p_3, p_4) = (p_{cc}, p_{cd}, p_{dc}, p_{dd})$, where p_z is the probability of playing c when the outcome z occurred in the previous round. If Y uses the memory-one plan $\mathbf{q} = (q_1, q_2, q_3, q_4)$, then the response vector is $(q_{cc}, q_{cd}, q_{dc}, q_{dd}) = (q_1, q_3, q_2, q_4)$ and the successive outcomes follow a Markov chain with transition matrix given by

$$\mathbf{M} = \begin{pmatrix} p_1q_1 & p_1(1-q_1) & (1-p_1)q_1 & (1-p_1)(1-q_1) \\ p_2q_3 & p_2(1-q_3) & (1-p_2)q_3 & (1-p_2)(1-q_3) \\ p_3q_2 & p_3(1-q_2) & (1-p_3)q_2 & (1-p_3)(1-q_2) \\ p_4q_4 & p_4(1-q_4) & (1-p_4)q_4 & (1-p_4)(1-q_4) \end{pmatrix}. \tag{1.7}$$

We use the switch in numbering from the Y strategy $\mathbf{q}$ to the Y response vector because switching the perspective of the players interchanges cd and dc. This way the "same" plan for X and for Y is given by the same vector. For example, TFT for X and for Y is given by $\mathbf{p} = \mathbf{q} = (1, 0, 1, 0)$, but the response vector for Y is $(1, 1, 0, 0)$. The plan *Repeat* is given by $\mathbf{p} = \mathbf{q} = (1, 1, 0, 0)$ with the response vector for Y equal to $(1, 0, 1, 0)$. This plan just repeats the previous play, regardless of what the opponent did.

We describe some elementary facts about finite Markov chains, see, e.g., Chapter 2 of Karlin and Taylor [10].

A Markov matrix like $\mathbf{M}$ is a non-negative matrix with row sums equal to 1. Thus, the vector $\mathbf{1}$ is a right eigenvector with eigenvalue 1. For such a matrix, we can represent the associated Markov chain as movement along a directed graph with vertices the states, in this case, cc, cd, dc, dd, and with a directed edge from the i-th state z_i to the j-th state z_j when $\mathbf{M}_{ij} > 0$, that is, when we can move from z_i to z_j with positive probability. In particular, there is an edge from z_i to itself iff the diagonal entry $\mathbf{M}_{ii}$ is positive.

A *path* in the graph is a state sequence $z^1, \dots, z^n$ with $n > 1$ such that there is an edge from z^i to z^{i+1} for $i = 1, \dots, n-1$. A set of states I is called a *closed set* when no path that begins in I can exit I. For example, the entire set of states is closed and for any z the set of states accessible via a path that begins at z is a closed set. I is closed iff $\mathbf{M}_{ij} = 0$ whenever $z_i \in I$ and $z_j \notin I$. In particular, when we restrict the chain to a closed set I, the associated submatrix of $\mathbf{M}$ still has row sums equal to 1. A minimal, nonempty, closed set of states is called a *terminal set*. A state is called *recurrent* when it lies in some terminal set and *transient* when it does not. The following facts are easy to check.

- A nonempty, closed set of states I is terminal iff for all $z_i, z_j \in I$, there exists a path from z_i to z_j.
- If I is a terminal set and $z_j \in I$, then there exists $z_i \in I$ with an edge from z_i to z_j.
- Distinct terminal sets are disjoint.
- Any nonempty, closed set contains at least one terminal set.
- From any transient state, there is a path into some terminal set.

Suppose we are given an initial distribution $\mathbf{v}^1$, describing the outcome of the first round of play. The Markov process evolves in discrete time via the equation

$$\mathbf{v}^{n+1} = \mathbf{v}^n \cdot \mathbf{M}, \tag{1.8}$$

where we regard the distributions as row vectors.

In our game context, the initial distribution is given by the initial plays, pure or mixed, of the two players. If X uses initial probability p_c and Y uses q_c, then

$$\mathbf{v}^1 = (p_c q_c, p_c(1-q_c), (1-p_c)q_c, (1-p_c)(1-q_c)). \tag{1.9}$$

Thus, v_i^n is the probability that outcome z_i occurs on the n-th round of play. A distribution $\mathbf{v}$ is *stationary* when it satisfies $\mathbf{vM} = \mathbf{v}$. That is, it is a left eigenvector with eigenvalue 1. From Perron–Frobenius theory (see, e.g., Appendix 2 of [10]), it follows that if I is a terminal set, then there is a unique stationary distribution $\mathbf{v}$ with $v_i > 0$ iff $i \in I$. That is, the *support* of $\mathbf{v}$ is exactly I. In particular, if the eigenspace of $\mathbf{M}$ associated with the eigenvalue 1 is one-dimensional, then there is a unique stationary distribution, and so a unique terminal set that is the support of the stationary distribution. The converse is also true and any stationary distribution $\mathbf{v}$ is a mixture of the $\mathbf{v}_J$'s, where $\mathbf{v}_J$ is supported on the terminal set J. This follows from the fact that any stationary distribution $\mathbf{v}$ satisfies $v_i = 0$ for all transient states z_i and so is supported on the set of recurrent states. On the recurrent states, the matrix $\mathbf{M}$ is block diagonal. Hence, the following are equivalent in our 4×4 case:

- There is a unique terminal set of states for the process associated with M.
- There is a unique stationary distribution vector for M.
- The matrix $M' = M - I$ has rank 3.

We will call $\mathbf{M}$ *convergent* when these conditions hold. For example, when all of the probabilities of $\mathbf{p}$ and $\mathbf{q}$ lie strictly between 0 and 1, then all the entries of $\mathbf{M}$ given by Equation (1.7) are positive and so the entire set of states is the unique terminal state and the positive matrix $\mathbf{M}$ is convergent.

The sequence of the Cesaro averages $\{1/n\Sigma_{i=1}^{n}\mathbf{v}^i\}$ of the outcome distributions always converges to some stationary distribution $\mathbf{v}$. That is,

$$\lim_{n\to\infty} \frac{1}{n}\Sigma_{k=1}^{n}\mathbf{v}^k = \mathbf{v}. \tag{1.10}$$

Hence, using the payoff vectors from Equation (1.2), the long-run average payoffs for X and Y converge to s_{X} and s_{Y} of Equation (1.3) with $\mathbf{v}$ this limiting stationary distribution.

When $\mathbf{M}$ is convergent, the limit $\mathbf{v}$ is the unique stationary distribution and so the average payoffs are independent of the initial distribution. In the non-convergent case, the long-term payoffs depend on the initial distribution. Suppose there are exactly two terminal sets, I and J, with stationary distribution vectors $\mathbf{v}_I$ and $\mathbf{v}_J$, supported on I and J, respectively. For any initial distribution $\mathbf{v}^1$, there are probabilities p_I and $p_J = 1 - p_I$ of entering into, and so terminating in, I or J, respectively. In that case, the limit of the Cesaro averages sequence for $\{\mathbf{v}^n\}$ is given by

$$\mathbf{v} = p_I\mathbf{v}_I + p_J\mathbf{v}_J, \tag{1.11}$$

and the limits of the average payoffs are given by Equation (1.3) with this distribution $\mathbf{v}$. This extends in the obvious way when there are more terminal sets.

When Y responds to the memory-one plan $\mathbf{p}$ with a memory-one plan $\mathbf{q}$, we have the *Markov case* as above. We will also want to see how a memory-one plan $\mathbf{p}$ for X fares against a not necessarily memory-one response by Y. We will call such a response pattern a *general plan* to emphasize that it need not be memory-one. That is, a general plan is a choice of response, pure or mixed, for any sequence of previous outcomes. Hereafter, unless we use the expression "general plan," we will assume a plan is memory-one.

If Y uses a general plan, then the sequence of Cesaro averages need not converge. We will call any limit point of the sequence an associated *limit distribution*. We will call s_{X} and s_{Y}, given by Equation (1.3) with such a limit distribution $\mathbf{v}$, the *expected payoffs* associated with $\mathbf{v}$.

Call a plan $\mathbf{p}$ *agreeable* when $p_1 = 1$ and *firm* when $p_4 = 0$. That is, an agreeable plan always responds to *cc* with *c* and a firm plan always responds to *dd* with *d*. If both $\mathbf{p}$ and $\mathbf{q}$ are agreeable, then $\{cc\}$ is a terminal set for the Markov matrix $\mathbf{M}$ given by Equation (1.7) and so $\mathbf{v} = (1, 0, 0, 0)$ is a stationary distribution with fixation at *cc*. If both $\mathbf{p}$ and $\mathbf{q}$ are firm, then $\{dd\}$ is a terminal set for $\mathbf{M}$ and $\mathbf{v} = (0, 0, 0, 1)$ is a stationary distribution with fixation at *dd*. Any convex combination of agreeable plans (or firm plans) is agreeable (respectively, firm).

An agreeable plan together with initial cooperation is called an *agreeable strategy*.

The plans $\mathrm{TFT} = (1, 0, 1, 0)$ and $\mathrm{Repeat} = (1, 1, 0, 0)$ are each agreeable and firm. The same is true for any mixture of these. If both X and Y use TFT, then the outcome is determined by the initial play. Initial outcomes *cc* and *dd* lead to immediate fixation. Either *cd* or *dc* results in period 2 alternation between these two states. Thus, $\{cd, dc\}$ is another terminal set with stationary distribution $(0, 1/2, 1/2, 0)$. If $a \cdot \mathrm{TFT} + (1 - a)\mathrm{Repeat}$ is used instead by either player (with $0 < a < 1$), then eventually fixation at *cc* or *dd* results. There are then only two terminal sets instead of three. The period 2 alternation described above illustrates why we needed the Cesaro limit, i.e., the limit of averages, in Equation (1.10) rather than the limit per se.

Because so much work had been done on this Markov model, the exciting new ideas in Press and Dyson [15] took people by surprise. They have inspired a number of responses, e.g., Stewart and Plotkin [16] and especially, Hilbe, Nowak, and Sigmund [8]. I would here like to express my gratitude to Karl Sigmund whose kind, but firm, criticism of the initial draft directed me to this recent work. The result is both a substantive and expository improvement.

Our purpose here is to use these new ideas to characterize the plans that are good in the following sense.

Definition 1.2. A plan $\mathbf{p}$ for X is called *good* if it is agreeable and if for any general plan chosen by Y against it and any associated limit distribution, the expected payoffs satisfy

$$s_Y \geq R \implies s_Y = s_X = R. \tag{1.12}$$

The plan is called *of Nash type* if it is agreeable and if the expected payoffs against any Y general plan satisfy

$$s_Y \geq R \implies s_Y = R. \tag{1.13}$$

A *good strategy* is a good plan together with initial cooperation.

By Proposition 1.1, $s_Y = s_X = R$ iff the associated limit distribution is $(1, 0, 0, 0)$. In the memory-one case, $(1, 0, 0, 0)$ is a stationary distribution iff both plans are agreeable. It is the unique stationary distribution iff, in addition, the matrix $\mathbf{M}$ is convergent. If $\mathbf{p}$ is not agreeable, then Equation (1.12) can be vacuously true. For example, if X plays All$D = (0, 0, 0, 0)$, then for any Y response $P \geq s_Y$ and the implication is true.

When both players use agreeable strategies, i.e., agreeable plans with initial cooperation, then the joint cooperative payoff is achieved. The pair of strategies is a Nash equilibrium exactly when the two plans are of Nash type. That is, both players receive R and neither player can do better by playing an alternative plan. A good plan is of Nash type, but more is true. We will see that with a Nash equilibrium it is possible that Y can play an alternative that still yields R for herself but with the payoff to X smaller than R. That is, Y has no incentive to play so as to reach the joint cooperative payoff. On the other hand, if X uses a good plan, then the only responses for Y that obtain R for her also yield R for X.

The strategy Repeat $= (1, 1, 0, 0)$ is an agreeable plan that is not of Nash type. If both players use Repeat, then the initial outcome repeats forever. If the initial outcome is cd, then $s_Y = T$ and $s_X = S$.

For a plan $\mathbf{p}$, we define the X *Press–Dyson vector* $\tilde{\mathbf{p}} = \mathbf{p} - \mathbf{e}_{12}$, where $\mathbf{e}_{12} = (1, 1, 0, 0)$. Considering the utility of the following result of Hilbe, Nowak and Sigmund, its proof, taken from Appendix A of [8], is remarkably simple.

Theorem 1.3. *Assume that* X *uses the plan* $\mathbf{p}$ *with* X *Press–Dyson vector* $\tilde{\mathbf{p}}$. *If the initial plays and the general plan of* Y *yields the sequence of distributions* $\{\mathbf{v}^n\}$, *then*

$$\begin{aligned} &\lim_{n\to\infty} \frac{1}{n} \Sigma_{k=1}^{n} \langle \mathbf{v}^k \cdot \tilde{\mathbf{p}} \rangle = 0, \\ &\textit{and so} \quad \langle \mathbf{v} \cdot \tilde{\mathbf{p}} \rangle = v_1\tilde{p}_1 + v_2\tilde{p}_2 + v_3\tilde{p}_3 + v_4\tilde{p}_4 = 0 \end{aligned} \tag{1.14}$$

for any associated limit distribution $\mathbf{v}$.

Proof. Let $v_{12}^n = v_1^n + v_2^n$, the probability that either cc or cd is the outcome in the n-th round of play. That is, $v_{12}^n = \langle \mathbf{v}^n \cdot \mathbf{e}_{12} \rangle$ is the probability that X played c in the n-th round. On the other hand, since X is using the plan $\mathbf{p}$, p_i is the conditional probability that X plays c in the next round, given outcome z_i in the current round. Thus, $\langle \mathbf{v}^n \cdot \mathbf{p} \rangle$ is the probability that X plays c in the $(n+1)^{st}$ round, i.e., it is v_{12}^{n+1}. Hence, $v_{12}^{n+1} - v_{12}^n = \langle \mathbf{v}^n \cdot \tilde{\mathbf{p}} \rangle$. The sum telescopes to yield

$$v_{12}^{n+1} - v_{12}^1 = \Sigma_{k=1}^{n} \langle \mathbf{v}^k \cdot \tilde{\mathbf{p}} \rangle. \tag{1.15}$$

As the left side has absolute value at most 1, Limit (1.14) follows. If a subsequence of the Cesaro averages converges to $\mathbf{v}$, then $\langle \mathbf{v} \cdot \tilde{\mathbf{p}} \rangle = 0$ by continuity of the dot product. □

To illustrate the use of this result, we examine $TFT = (1, 0, 1, 0)$ and another plan that has been labeled in the literature Grim $= (1, 0, 0, 0)$. We consider mixtures of each with Repeat $= (1, 1, 0, 0)$.

Corollary 1.4. *Let* $1 \geq a > 0$.

(a) *The plan* $\mathbf{p} = a\mathrm{TFT} + (1-a)\mathrm{Repeat}$ *is a good plan with* $s_Y = s_X$ *for any limiting distribution.*

(b) *The plan* $\mathbf{p} = a\mathrm{Grim} + (1-a)\mathrm{Repeat}$ *is good.*

Proof. (a) In this case, $\tilde{\mathbf{p}} = a(0,-1,1,0)$ and so Equation (1.14) implies that $v_2 = v_3$. Thus, $s_Y = s_X$. From this Equation (1.12) follows from Proposition 1.1.

(b) Now $\tilde{\mathbf{p}} = a(0,-1,0,0)$ and so Equation (1.14) implies that $v_2 = 0$. Thus, $s_Y = v_1R + v_3S + v_4P$ and this is less than R unless $v_3 = v_4 = 0$ and $v_1 = 1$. When $v_1 = 1$, $s_Y = s_X = R$, proving Equation (1.12). □

In the next section, we will prove the following characterization of the good plans.

Theorem 1.5. *Let* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *be an agreeable plan other than* Repeat. *That is,* $p_1 = 1$ *but* $\mathbf{p} \neq (1,1,0,0)$.

The plan $\mathbf{p}$ *is of Nash type iff the following inequalities hold.*

$$\frac{T-R}{R-S} \cdot p_3 \leq (1-p_2) \quad \text{and} \quad \frac{T-R}{R-P} \cdot p_4 \leq (1-p_2). \tag{1.16}$$

The plan $\mathbf{p}$ *is good iff, in addition, both inequalities are strict.*

Corollary 1.6. *In the compact convex set of agreeable plans, the set* $\{\mathbf{p}$ *equals* Repeat *or is of Nash type*$\}$ *is a closed convex set with interior set of good plans.*

Proof. The X Press–Dyson vectors form a cube and the agreeable plans are the intersection with the subspace $\tilde{p}_1 = 0$. We then intersect with the half-spaces defined by

$$\frac{T-R}{R-S}\tilde{p}_3 + \tilde{p}_2 \leq 0 \quad \text{and} \quad \frac{T-R}{R-P}\tilde{p}_4 + \tilde{p}_2 \leq 0. \tag{1.17}$$

The result is a closed convex set with interior given by the strict inequalities. Notice that these conditions are preserved by multiplication by a positive constant $a \leq 1$ or by any larger constant so long as $a\tilde{\mathbf{p}}$ remains in the cube. Hence, Repeat with $\tilde{\mathbf{p}} = 0$ is on the boundary. □

It is easy to compute that

$$\det \begin{pmatrix} R & R & 1 & 0 \\ S & T & 1 & 1 \\ T & S & 1 & 1 \\ P & P & 1 & 0 \end{pmatrix} = -2(R-P)(T-S). \tag{1.18}$$

Hence, with $\mathbf{e}_{23} = (0,1,1,0)$, we can use $\{\mathbf{S}_X, \mathbf{S}_Y, \mathbf{1}, \mathbf{e}_{23}\}$ as a basis for $\mathbb{R}^4$. For a distribution vector $\mathbf{v}$, we will write v_{23} for $v_2 + v_3 = \langle \mathbf{v} \cdot \mathbf{e}_{23} \rangle$. From Theorem 1.3, we immediately obtain the following.

Theorem 1.7. *If* $\mathbf{p}$ *is a plan whose X Press–Dyson vector* $\tilde{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$ *and* $\mathbf{v}$ *is a limit distribution when Y plays some general plan against* $\mathbf{p}$*, then the average payoffs satisfy the following* Press–Dyson equation:

$$\alpha s_X + \beta s_Y + \gamma + \delta v_{23} = 0. \tag{1.19}$$

The most convenient cases to study occur when $\delta = 0$. Press and Dyson called such a plan a *zero-determinant strategy* (hereafter ZDS) because of an ingenious determinant argument leading to Equation (1.19). We have used Theorem 1.3 of Hilbe–Nowak–Sigmund instead.

This representation yields a simple description of the good plans.

Theorem 1.8. *Assume that* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *is an agreeable plan with* X *Press–Dyson vector* $\tilde{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$*. Assume that* $\mathbf{p}$ *is not* Repeat, *i.e.*, $(\alpha, \beta, \gamma, \delta) \neq (0, 0, 0, 0)$*. The plan* $\mathbf{p}$ *is of Nash type iff*

$$\max\left(\frac{\delta}{(T-S)}, \frac{\delta}{(2R-(T+S))}\right) \leq \alpha. \tag{1.20}$$

The plan $\mathbf{p}$ *is good iff, in addition, the inequality is strict.*

Remark: Observe that $T - S > 2R - (T + S) > 0$. It follows that if $\delta \leq 0$, then $\mathbf{p}$ is good iff $\delta/(T - S) < \alpha$. On the other hand, if $\delta > 0$, then $\mathbf{p}$ is good iff $\delta/(2R - (T + S)) < \alpha$.

In the next section, we will investigate the geometry of the $\{\mathbf{S}_X, \mathbf{S}_Y, \mathbf{1}, \mathbf{e}_{23}\}$ decomposition of the Press–Dyson vectors and prove the theorems.

2 Good plans and the Press–Dyson decomposition

We begin by normalizing the payoffs. We can add to all a common number and multiply all by a common positive number without changing the relationship between the various strategies. We subtract S and divide by $T - S$. So from now on we will assume that $T = 1$ and $S = 0$.

The payoff vectors of Equation (1.2) are then given by

$$\mathbf{S}_X = (R, 0, 1, P), \quad \mathbf{S}_Y = (R, 1, 0, P), \tag{2.1}$$

and from Equation (1.4), we have

$$1 > R > \frac{1}{2}, \quad \text{and} \quad R > P > 0. \tag{2.2}$$

After normalization Theorem 1.5 becomes the following.

Theorem 2.1. *Let* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *be an agreeable plan other than* Repeat. *That is,* $p_1 = 1$ *but* $\mathbf{p} \neq (1, 1, 0, 0)$.

The plan **p** *is of Nash type iff the following inequalities hold.*

$$\frac{1-R}{R} \cdot p_3 \le (1-p_2) \quad \textit{and} \quad \frac{1-R}{R-P} \cdot p_4 \le (1-p_2). \tag{2.3}$$

The plan **p** *is good iff, in addition, both inequalities are strict.*

Proof. We first eliminate the possibility $p_2 = 1$. If $1 - p_2 = 0$, then the inequalities would yield $p_3 = p_4 = 0$ and so $\mathbf{p} =$ Repeat, which we have excluded. On the other hand, if $p_2 = 1$, then $\mathbf{p} = (1, 1, p_3, p_4)$. If against this Y plays All$D = (0, 0, 0, 0)$, then $\{cd\}$ is a terminal set with stationary distribution $(0, 1, 0, 0)$ and so with $s_Y = 1$ and $s_X = 0$. Hence, **p** is not of Nash type. Thus, if $p_2 = 1$, then neither is **p** of Nash type nor do the inequalities hold for it. We now assume $1 - p_2 > 0$.

Observe that

$$\begin{aligned} s_Y - R &= (v_1 R + v_2 + v_4 P) - (v_1 R + v_2 R + v_3 R + v_4 R) \\ &= v_2(1-R) - v_3 R - v_4(R-P). \end{aligned} \tag{2.4}$$

Hence, multiplying by the positive quantity $(1 - p_2)$, we have

$$s_Y >= R \iff (1-p_2)v_2(1-R) >= v_3(1-p_2)R + v_4(1-p_2)(R-P), \tag{2.5}$$

where this notation means that the inequalities are equivalent and the equations are equivalent.

Since $\tilde{p}_1 = 0$, Equation (1.14) implies $v_2\tilde{p}_2 + v_3\tilde{p}_3 + v_4\tilde{p}_4 = 0$ and so $(1-p_2)v_2 = v_3 p_3 + v_4 p_4$. Substituting in the above inequality and collecting terms we get

$$\begin{aligned} &s_Y >= R \iff Av_3 >= Bv_4, \text{ with} \\ &A = [p_3(1-R) - (1-p_2)R] \quad \text{and} \quad B = [(1-p_2)(R-P) - p_4(1-R)]. \end{aligned} \tag{2.6}$$

Observe that the inequalities of Equation (2.3) are equivalent to $A \le 0$ and $B \ge 0$. The proof is completed by using a sequence of little cases.

Case (i) $A = 0, B = 0$: In this case, $Av_3 = Bv_4$ holds for any strategy for Y. So for any Y strategy, $s_Y = R$ and **p** is of Nash type. If Y chooses a plan that is not agreeable, then $\{cc\}$ is not a closed set of states and so $v_1 \ne 1$. From Proposition 1.1, $s_X < R$ and so **p** is not good.

Case (ii) $A < 0, B = 0$: The inequality $Av_3 \ge Bv_4$ holds iff $v_3 = 0$. If $v_3 = 0$, then $Av_3 = Bv_4$ and so $s_Y = R$. Thus, **p** is Nash.

Case (iia) $B \le 0$, any A: Assume Y chooses a plan that is not agreeable and is such that $v_3 = 0$. For example, if Y plays All$D = (0, 0, 0, 0)$, then no state moves to dc. With such a Y choice, $Av_3 \ge Bv_4$ and so $s_Y \ge R$. As above, $v_1 \ne 1$ because the Y plan is not agreeable. Again $s_X < R$ and **p** is not good. Furthermore, $v_3 = 0$, $v_1 < 1$, $p_2 < 1$, and $(1-p_2)v_2 = v_4 p_4$ imply that $v_4 > 0$. So if $B < 0$, then $Av_3 > Bv_4$ and so $s_Y > R$. Hence, **p** is not Nash when $B < 0$.

Case (iii) $A = 0, B > 0$: The inequality $Av_3 \ge Bv_4$ holds iff $v_4 = 0$. If $v_4 = 0$, then $Av_3 = Bv_4$ and $s_Y = R$. Thus, **p** is Nash.

Case (iiia) $A \geq 0$, any B: Assume Y chooses a plan that is not agreeable and is such that $v_4 = 0$. For example, if Y plays $(0, 1, 1, 1)$, then no state moves to dd. With such a Y choice, $Av_3 \geq Bv_4$ and so $s_Y \geq R$. As before, $v_1 \neq 1$ implies $s_X < R$ and the plan is not good. Furthermore, $v_4 = 0$, $v_1 < 1$, $p_2 < 1$, and $(1 - p_2)v_2 = v_3p_3$ imply that $v_3 > 0$. So if $A > 0$, then $Av_3 > Bv_4$ and so $s_Y > R$. Hence, $\mathbf{p}$ is not Nash when $A > 0$.

Case (iv) $A < 0$, $B > 0$: The inequality $Av_3 \geq Bv_4$ implies $v_3, v_4 = 0$. So $(1 - p_2)v_2 = v_3p_3 + v_4p_4 = 0$. Since $p_2 < 1$, $v_2 = 0$. Hence, $v_1 = 1$. That is, $s_Y \geq R$ implies $s_Y = s_X = R$ and so $\mathbf{p}$ is good. □

Remarks:

(a) Since $1 > R > 1/2$, it is always true that $(1 - R)/R < 1$. On the other hand, $(1 - R)/(R - P)$ can be greater than 1 and the second inequality requires $p_4 \leq (R - P)/(1 - R)$. In particular, if $p_2 = 0$, then the plan is good iff $p_4 < (R - P)/(1 - R)$. For example, the plan $(1, 0, 0, 1)$ is, in the literature, labeled *Pavlov*, or *WinStay, LoseShift*. This plan always satisfies the first inequality strictly, but it satisfies the second strictly, and so is good, iff $1 - R < R - P$.

(b) In Case (i) of the proof, the payoff $s_Y = R$ is determined by $\mathbf{p}$ independent of the choice of strategy for Y. In general, plans that fix the opponent's payoff in this way were described by Press and Dyson [15] and, earlier, by Boerlijst, Nowak, and Sigmund [5], where they are called *equalizer strategies*. The agreeable equalizer plans have $\tilde{\mathbf{p}} = a(0, -(1 - R)/R, 1, (R - P)/R)$ with $1 \geq a > 0$.

Christian Hilbe suggests a nice interpretation of the above results.

Corollary 2.2. *Let* $\mathbf{p}$ *be an agreeable plan with* $p_2 < 1$.

(a) *If* $\mathbf{p}$ *is good, then using any plan* $\mathbf{q}$ *that is not agreeable forces* Y *to get a payoff* $s_Y < R$.

(b) *If* $\mathbf{p}$ *is not good, then by using at least one of the two plans* $\mathbf{q} = (0, 0, 0, 0)$ *or* $\mathbf{q} = (0, 1, 1, 1)$, Y *can certainly obtain a payoff* $s_Y \geq R$*, and force* X *to get a payoff* $s_X < R$.

(c) *If* $\mathbf{p}$ *is not Nash, then by using at least one of the two plans* $\mathbf{q} = (0, 0, 0, 0)$ *or* $\mathbf{q} = (0, 1, 1, 1)$*, Y can certainly obtain a payoff* $s_Y > R$*, and force* X *to get a payoff* $s_X < R$.

Proof. (a) If $\mathbf{p}$ is good, then $s_Y \geq R$ implies $s_Y = s_X = R$, which requires $v = (1, 0, 0, 0)$. This is only stationary when $\mathbf{q}$ as well as $\mathbf{p}$ is agreeable.
(b) and (c) follow from the analysis of cases in the above proof. □

Remark: If $p_2 = p_1 = 1$, then the plan $\mathbf{p}$ is not Nash. As observed in the proof of Theorem 2.1, if Y plays $\mathbf{q} = (0, 0, 0, 0)$, then cd is a terminal set with stationary distribution

$\mathbf{v} = (0, 1, 0, 0)$ and so with $s_Y = 1$, $s_X = 0$. However, if, in addition, $p_4 = 0$, e.g., if X uses Repeat, then *dd* is also a terminal set. Thus, if X plays $\mathbf{p}$ with $1-p_4 = p_2 = p_1 = 1$ and Y always defects, then fixation occurs immediately at either *cd* with $s_Y = 1$ and $s_X = 0$, or else at *dd* with $s_Y = s_X = P$. The result is determined by the initial play of X.

We now consider the Press–Dyson representation, using the normalized payoff vectors of Equation (2.1). If $\tilde{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$ is the X Press–Dyson vector of a plan $\mathbf{p}$, then it must satisfy two sorts of constraints.

The *sign constraints* require that the first two entries be non-positive and the last two be non-negative. That is,

$$\begin{aligned} &(\alpha+\beta)R + \gamma \le 0, \\ &\beta + \gamma + \delta \le 0, \\ &\alpha + \gamma + \delta \ge 0, \\ &(\alpha+\beta)P + \gamma \ge 0. \end{aligned} \tag{2.7}$$

Lemma 2.3. *If* $\tilde{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$ *satisfies the sign constraints, then*

$$\begin{aligned} &\alpha + \beta \le 0 \quad \textit{and} \quad \gamma \ge 0, \\ &\alpha + \beta = 0 \ \Leftrightarrow \ \gamma = 0. \end{aligned} \tag{2.8}$$

Proof. Subtracting the fourth inequality from the first we see that $(\alpha+\beta)(R-P) \le 0$ and so $R - P > 0$ implies $\alpha + \beta \le 0$. Then the fourth inequality and $P > 0$ imply $\gamma \ge 0$. The first and fourth imply $\alpha + \beta = 0$ iff $\gamma = 0$. □

Remark: Notice that both $\tilde{p}_1$ and $\tilde{p}_4$ vanish iff $\alpha + \beta = \gamma = 0$. These are the cases when plan $\mathbf{p}$ is both agreeable and firm.

In addition, the entries of an X Press–Dyson vector have absolute value at most 1. These are the *size constraints*. If a vector satisfies the sign constraints, then, multiplying by a sufficiently small positive number, we obtain the size constraints as well. Any vector in $\mathbb{R}^4$ that satisfies both the sign and the size constraints is an X Press–Dyson vector. Call $\mathbf{p}$ a *top plan* if $|\tilde{p}_i| = 1$ for some i. For any plan $\mathbf{p}$, other than Repeat, which has X Press–Dyson vector $\mathbf{0}$, $\mathbf{p} = a(\mathbf{p}^t) + (1-a)$Repeat for a unique top plan $\mathbf{p}^t$ and a unique positive $a \le 1$. Equivalently, $\tilde{\mathbf{p}} = a\tilde{\mathbf{p}}^t$.

Observe that $\mathbf{p}$ is agreeable iff $\tilde{p}_1 = 0$ and so iff $(\alpha+\beta)R + \gamma = 0$. In that case, $\beta = -\alpha - \gamma R^{-1}$. Substituting into Equation (1.19), we obtain the following corollary of Theorem 1.7.

Corollary 2.4. *If* $\mathbf{p}$ *is an agreeable plan with* X *Press–Dyson vector* $\tilde{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$, *then the payoffs with any limit distribution satisfy the following version of the* Press–Dyson equation.

$$\gamma R^{-1} s_Y + \alpha(s_Y - s_X) - \delta v_{23} = \gamma. \tag{2.9}$$

Now we justify the description in Theorem 1.8. Notice that if we label by S_X^0 and S_Y^0 our original payoff vectors before normalization, then $\mathbf{S}_X^0 = (T-S)\mathbf{S}_X + S\mathbf{1}$, $S_Y^0 = (T-S)\mathbf{S}_Y + S\mathbf{1}$ and so if $(\alpha, \beta, \gamma, \delta)$ are the coordinates of $\tilde{\mathbf{p}}$ with respect to the basis $\{\mathbf{S}_X, \mathbf{S}_Y, \mathbf{1}, \mathbf{e}_{23}\}$, then $(\alpha^0, \beta^0, \gamma^0, \delta^0) = (\alpha/(T-S), \beta/(T-S), \gamma - (\alpha+\beta)S/(T-S), \delta)$ are the coordinates with respect to $\{S_X^0, S_Y^0, \mathbf{1}, \mathbf{e}_{23}\}$. In particular, $\alpha >= k\delta$ iff $\alpha^0 >= k\delta^0/(T-S)$ for any k. Furthermore, the constant $k = (T-S)/(2R-(T+S))$ is independent of normalization. So it suffices to prove the normalized version of the theorem, which is the following.

Theorem 2.5. *Assume that* $\mathbf{p} = (p_1, p_2, p_3, p_4)$ *is an agreeable plan with X Press–Dyson vector* $\tilde{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$. *Assume that* $\mathbf{p}$ *is not* Repeat, *i.e.*, $(\alpha, \beta, \gamma, \delta) \neq (0,0,0,0)$. *The plan* $\mathbf{p}$ *is of Nash type iff*

$$\max(\delta, (2R-1)^{-1}\delta) \leq \alpha. \tag{2.10}$$

The plan $\mathbf{p}$ *is good iff, in addition, the inequality is strict.*

Proof. Since $\beta = -\alpha - \gamma R^{-1}$, we have

$$\begin{aligned} (1-p_2) &= -\tilde{p}_2 = -\beta - \gamma - \delta = \alpha + \frac{1-R}{R}\gamma - \delta, \\ p_3 = \tilde{p}_3 &= \alpha + \gamma + \delta, \quad p_4 = \tilde{p}_4 = \frac{R-P}{R}\gamma. \end{aligned} \tag{2.11}$$

The inequality $(1-R)p_3 \leq R(1-p_2)$ becomes $(1-R)(\alpha+\gamma+\delta) \leq R\alpha + (1-R)\gamma - R\delta$. This reduces to $\delta \leq (2R-1)\alpha$. Similarly, the inequality $(1-R)p_4 \leq (R-P)(1-p_2)$ reduces to $\delta \leq \alpha$. $\square$

Remarks: (a) Thus, when $\delta \leq 0$, $\mathbf{p}$ is good iff $\delta < \alpha$. When $\delta > 0$, $\mathbf{p}$ is good iff $\delta/(2R-1) < \alpha$.

(b) From the proof we see that the equalizer case, when both inequalities of Equation (2.3) are equations, occurs when $\delta = \alpha = (2R-1)^{-1}\delta$. Since $2R-1 < 1$, this reduces to $0 = \delta = \alpha$.

In the ZDS case, when $\delta = 0$, we can rewrite Equation (2.9) as

$$\kappa \cdot (s_X - R) = s_Y - R \tag{2.12}$$

with $\kappa = \alpha R/(\gamma + \alpha R)$. Thus, the condition $\alpha > 0$ is equivalent to $0 < \kappa \leq 1$. In [8], these plans are introduced and called *complier strategies*. The equation and the condition $\kappa > 0$ make it clear that such plans are good. In addition, if $s_Y < R$, then it follows that $s_X \leq s_Y$ with strict inequality when $\gamma > 0$ and so $\kappa < 1$. The strategy ZGTFT-2 analyzed in Stewart and Plotkin [16] is an example of a complier plan. When X plays a complier plan, then either both s_X and s_Y are equal to R or else both are below R. This is not true for good plans in general. If X plays the good plan Grim $= (1,0,0,0)$ and Y plays $(0,1,1,1)$, then fixation at *dc* occurs with $v = (0,0,1,0)$ and so with $s_Y = 0$ ($<$R as required by Corollary 2.2), but with $s_X = 1 > R$.

Let us look at the geometry of the Press–Dyson representation.

We begin with the *exceptional plans* that are defined by $\gamma = \alpha + \beta = 0$. The sign constraints yield $\alpha = -\beta \geq |\delta|$ and $\tilde{\mathbf{p}} = (0, \delta - \alpha, \delta + \alpha, 0)$. As remarked after Lemma 2.3, the exceptional plans are exactly those plans that are both agreeable and firm. In the xy plane with $x = \tilde{p}_2$ and $y = \tilde{p}_3$, they form a square with vertices: Repeat($\tilde{\mathbf{p}} = (0,0,0,0)$), Grim($\tilde{\mathbf{p}} = (0,-1,0,0)$), TFT ($\tilde{\mathbf{p}} = (0,-1,1,0)$), and what we will call Lame ($\tilde{\mathbf{p}} = (0,0,1,0)$).

Thus, Lame $= (1,1,1,0)$. The top plans consist of the segment that connects TFT with Grim together with the segment that connects TFT with Lame.

On the Grim-TFT segment, $\delta - \alpha = -1$ and $0 \leq \delta + \alpha \leq 1$. That is, $\delta = \alpha - 1$ and $-1/2 \leq \delta \leq 0$. By Theorem 2.5, the plans in the triangle with vertices Grim, TFT, and Repeat are all good except for Repeat itself.

On the Lame-TFT segment, $-1 \leq \delta - \alpha \leq 0$ and $\delta + \alpha = 1$. That is, $\delta = 1 - \alpha$ and $1/2 \geq \delta \geq 0$. Such a plan is the mixture tTFT $+ (1-t)$Lame $= (1, 1-t, 1, 0)$ with $t = 2\alpha - 1$. By Theorem 2.1, this plan is good iff $t > (1-R)/R$. The plan on the TFT-Lame segment with $t = (1-R)/R$, and so with $2\alpha = R^{-1}$, we will call Edge $= (1, (2R-1)/R, 1, 0)$. The plans in the TFT-Edge-Repeat triangle that are not on the Edge-Repeat side are good plans. The plans in the complementary Edge-Lame-Repeat triangle are not good.

Now assume $\gamma > 0$, and define

$$\bar{\alpha} = \alpha/\gamma, \quad \bar{\beta} = \beta/\gamma, \quad \bar{\delta} = \delta/\gamma. \tag{2.13}$$

with the sign constraints

$$\begin{aligned} -P^{-1} &\leq \bar{\alpha} + \bar{\beta} \leq -R^{-1}, \\ \bar{\beta} &\leq -1 - \bar{\delta} \leq \bar{\alpha}. \end{aligned} \tag{2.14}$$

For any triple $(\bar{\alpha}, \bar{\beta}, \bar{\delta})$ that satisfies these inequalities, we obtain an X Press–Dyson vector $\tilde{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$ that satisfies the size constraints as well by using $(\alpha, \beta, \gamma, \delta) = \gamma \cdot (\bar{\alpha}, \bar{\beta}, 1, \bar{\delta})$ with $\gamma > 0$ small enough. When we use the largest value of γ such that the size constraints hold, we obtain the top plan associated with $(\bar{\alpha}, \bar{\beta}, \bar{\delta})$. The others are mixtures of the top plan with Repeat. For a plan with this triple, the Press–Dyson equation (1.19) becomes

$$\bar{\alpha}s_X + \bar{\beta}s_Y + \bar{\delta}v_{23} + 1 = 0. \tag{2.15}$$

The points $(x, y) = (\bar{\alpha}, \bar{\beta})$ lie in the *Strategy Strip*. This consists of the points of the xy plane with $y \leq x$ and that lie on or below the line $x + y = -R^{-1}$ and on or above the line $x + y = -P^{-1}$. Then $\bar{\delta}$ must satisfy $-1 - x \leq \bar{\delta} \leq -1 - y$. Alternatively, we can fix $\bar{\delta}$ to be arbitrary and intersect the Strategy Strip with the fourth quadrant when the origin is at $(-1 - \bar{\delta}, -1 - \bar{\delta})$, i.e., the points with $y \leq -1 - \bar{\delta} \leq x$.

Together with the exceptional plans those with $(\bar{\alpha}, \bar{\beta})$ on the line $x + y = -R^{-1}$ are exactly the agreeable plans. Together with the exceptional plans those on the line $x + y = -P^{-1}$ are exactly the firm plans.

Let us look at the good ZDS's, i.e., the good plans with $\delta = 0$. In the exceptional case with $\gamma = 0$, the top good plan is TFT. When $\delta = 0$ and $\gamma > 0$, the good plans are those that satisfy $\bar{\alpha} + \bar{\beta} = -R^{-1}$ and $\bar{\alpha} > 0$. As mentioned above, these are the complier plans.

Proposition 2.6. *Given $\bar{\alpha} > 0$, the associated agreeable ZDS top plan is given by*

$$\mathbf{p} = \left(1, \frac{2R-1}{R(\bar{\alpha}+1)}, 1, \frac{R-P}{R(\bar{\alpha}+1)}\right). \tag{2.16}$$

Proof. The agreeable plan $\mathbf{p}$ with $\gamma, \bar{\alpha} > 0$ and $\bar{\delta} = 0$ has X Press–Dyson vector

$$\tilde{\mathbf{p}} = \big(0, -\gamma(\bar{\alpha} + R^{-1} - 1), \gamma(\bar{\alpha}+1), \gamma(1 - P \cdot R^{-1})\big). \tag{2.17}$$

With $\bar{\alpha}$ fixed, the largest value for γ so that the size constraints hold is $(\bar{\alpha}+1)^{-1}$. This easily yields Equation (2.16) for the top plan. □

When $\bar{\delta} = 0$, the vertical line $\bar{\alpha} = 0$ intersects the strip in points whose plans are all the *equalizers*, as discussed by Press and Dyson [15] and by Boerlijst, Nowak, and Sigmund [5]. Observe that with $\bar{\delta} = 0$, and $\bar{\alpha} = 0$, the Press–Dyson equation (2.15) becomes $\bar{\beta} s_{\mathrm{Y}} + 1 = 0$, and so $s_{\mathrm{Y}} = -\bar{\beta}^{-1}$ regardless of the choice of strategy for Y. The agreeable case has $\bar{\beta} = -R^{-1}$. The vertical line of equalizers cuts the line of agreeable plans, separating it into the unbounded ray with good plans and the segment with plans that are not even of Nash type.

Finally, we call a plan $\mathbf{p}$ *generous* when $p_2 > 0$ and $p_4 > 0$. That is, whenever Y defects, there is a positive probability that X will cooperate. The complier plans given by Equation (2.16) are generous.

Proposition 2.7. *Assume that X plays $\mathbf{p}$, a generous plan of Nash type. If Y plays plan $\mathbf{q}$ of Nash type and either (i) $\mathbf{q}$ is generous or (ii) $q_3 + q_4 > 0$, then $\{cc\}$ is the unique terminal set for the associated Markov matrix $\mathbf{M}$. Thus, $\mathbf{M}$ is convergent.*

Proof. Since $\mathbf{p}$ and $\mathbf{q}$ are both agreeable, $\{cc\}$ is a terminal set for $\mathbf{M}$.

Since $\mathbf{p}$ is of Nash type, it is not Repeat and so Equation (2.3) implies that $p_2 < 1$.

For the first case, we prove that if $p_1 = 1, p_2 < 1, p_4 > 0$ and $\mathbf{q}$ satisfies analogous conditions and not both $\mathbf{p}$ and $\mathbf{q}$ are of the form $(1, 0, 1, a)$, then $\mathbf{M}$ is convergent.

Recall that Y responds to *cd* using q_3 and to *dc* using q_2.

The assumptions $p_4, q_4 > 0$ imply that there is an edge from *dd* to *cc*, and so that *dd* is transient. There is an edge from *dc* to *dd* if $p_3 < 1$ since $q_2 < 1$. If $p_3 = 1$ and $q_2 > 0$, then there is an edge to *cc*. There remains the case that $p_3 = 1, q_2 = 0$ with the only edge from *dc* going to *cd*. Similarly, there is an edge from *cd* to either *dd* or *cc* except when $p_2 = 0, q_3 = 1$. Thus, the only case when $\mathbf{M}$ is not convergent is when both $\mathbf{p}$ and $\mathbf{q}$ are of the form $(1, 0, 1, a)$. In that case, $\{cd, dc\}$ is an additional terminal set. In particular, if either p_2 or q_2 is positive, then $\{cc\}$ is the only terminal set. This completes case (i). It also shows that if $\mathbf{p}$ is generous and $q_4 > 0$, then $\mathbf{M}$ is convergent.

To complete case (ii), we assume that $\mathbf{p}$ is generous and $q_3 > 0$. Since $p_2 > 0$ and $q_3 > 0$, there is an edge from cd to cc and so cd is transient. Since $p_4 > 0$, there is an edge from dd either to cc or to cd and so dd is transient. Finally, $q_2 < 1$ implies there is an edge from dc to cd or to dd. Thus, dc is transient as well. □

This result indicates the advantage which the good plans that are generous have over the good exceptional plans like Grim and TFT. The latter are firm as well as agreeable. Playing them against each other yields a non-convergent matrix with both $\{cc\}$ and $\{dd\}$ as terminal sets. Initial cooperation does lead to immediate fixation at cc, but an error might move the sequence of outcomes on a path leading to another terminal set. When generous good plans are used against each other, $\{cc\}$ is the unique terminal set. Eventual fixation at cc occurs whatever the initial distribution is, and if an error occurs, then the strategies move the successive outcomes along a path that returns to cc. It is easy to compute the expected number of steps T_z from transient state z to cc.

$$T_z = 1 + \Sigma_{z'} p_{zz'} T_{z'}, \tag{2.18}$$

where we sum over the three transient states and $p_{zz'}$ is the probability of moving along an edge from z to z'. Thus, with $\mathbf{M}' = \mathbf{M} - I$, we obtain the formula for the vector $\mathbf{T} = (T_2, T_3, T_4)$:

$$\mathbf{M}'_t \cdot \mathbf{T} = -\mathbf{1}, \tag{2.19}$$

where $\mathbf{M}'_t$ is the invertible 3×3 matrix obtained from $\mathbf{M}'$ by omitting the first row and column.

Consider the case when X and Y both use the plan given by Equation (2.16), so that $\mathbf{p} = \mathbf{q} = (1, p_2, 1, p_4)$. The only edges coming from cd connect with cc or with dc and similarly for the edges from dc. Symmetry will imply that $T_{cd} = T_{dc}$. So with T this common value we obtain from Equation (2.18) $T = 1 + (1 - p_2)T$. Hence, from Equation (2.16), we get

$$T = T_{cd} = T_{dc} = \frac{1}{p_2} = \frac{\bar{\alpha} + 1}{2 - R^{-1}}. \tag{2.20}$$

Thus, the closer the plan is to the equalizer plan with $\bar{\alpha} = 0$, the shorter the expected recovery time from an error leading to a dc or cd outcome. From Equation (2.18), one can see that

$$T_{dd} = 1 + 2p_4(1 - p_4) \cdot T + (1 - p_4)^2 \cdot T_{dd}. \tag{2.21}$$

We won't examine this further as arriving at dd from cc implies errors on the part of both players.

Of course, one might regard such departures from cooperation not as noise or error but as ploys. Y might try a rare move to *cd* in order to pick up the temptation payoff for defection as an occasional bonus. But if this is strategy rather than error, it means that Y is departing from the good plan to one with q_1 a bit less than 1. Corollary 2.2(a) implies that Y loses by executing such a ploy.

3 Competing zero-determinant strategies

We now examine the ZDS's in more detail. Recall that a plan $\mathbf{p}$ is a ZDS when $\delta = 0$ in the Press–Dyson decomposition of the X Press–Dyson vector $\tilde{\mathbf{p}} = \mathbf{p} - \mathbf{e}_{12}$. With Normalization (2.1), the inverse matrix of $(\mathbf{S}_X\ \mathbf{S}_Y\ \mathbf{1}\ \mathbf{e}_{23})$ is

$$\frac{-1}{2(R-P)}\begin{pmatrix} -1 & R-P & P-R & 1 \\ -1 & P-R & R-P & 1 \\ 2P & 0 & 0 & -2R \\ 1-2P & P-R & P-R & 2R-1 \end{pmatrix} \tag{3.1}$$

and so if $\tilde{\mathbf{p}} = \alpha\mathbf{S}_X + \beta\mathbf{S}_Y + \gamma\mathbf{1} + \delta\mathbf{e}_{23}$,

$$2(R-P)\delta = (2P-1)\tilde{p}_1 + (R-P)(\tilde{p}_2 + \tilde{p}_3) - (2R-1)\tilde{p}_4. \tag{3.2}$$

Thus, for example, if $R + P = 1$, both All$D$ with $\tilde{\mathbf{p}} = (-1,-1,0,0)$ and All$C$ with $\tilde{\mathbf{p}} = (0,0,1,1)$ are ZDS.

The exceptional ZDSs, which have $\gamma = 0$ as well as $\delta = 0$, are mixtures of TFT and Repeat. Otherwise, $\gamma > 0$ and we can write $\tilde{\mathbf{p}} = \gamma(\bar{\alpha}\mathbf{S}_X + \bar{\beta}\mathbf{S}_Y + \mathbf{1})$. When $(\bar{\alpha},\bar{\beta})$ lies in the ZDS strip defined by

$$\text{ZDS strip} = \{(x,y) : x \geq -1 \geq y \quad \text{and} \quad -R^{-1} \geq x+y \geq -P^{-1}\}, \tag{3.3}$$

then the sign constraints are satisfied. The size constraints hold as well when $\gamma > 0$ is small enough. For Z with $P \leq Z \leq R$ the intersection of the ZDS strip with the line $x + y = -Z^{-1}$ is a *value line* in the strip.

Lemma 3.1. *Assume that $(\bar{\alpha},\bar{\beta})$ in the ZDS strip, with $\bar{\alpha} + \bar{\beta} = -Z^{-1}$. We then have $-\bar{\beta} \geq \max(1,|\bar{\alpha}|)$ and $-\bar{\beta} = |\bar{\alpha}|$ iff $\bar{\alpha} = \bar{\beta} = -1$. If $(\bar{a},\bar{b})$ is also in the strip, then $D = \bar{\beta}\bar{b} - \bar{\alpha}\bar{a} \geq 0$ with equality iff $\bar{\alpha} = \bar{\beta} = \bar{a} = \bar{b} = -1$.*

Proof. By definition of Z, $-\bar{\beta} = \bar{\alpha} + Z^{-1} > \bar{\alpha}$. Also, the sign constraints imply $-\bar{\beta} \geq 1 \geq -\bar{\alpha}$, and so $-\bar{\beta} \geq -\bar{\alpha}$ with equality iff $\bar{\alpha} = \bar{\beta} = -1$. $D \geq (-\bar{\beta})(-\bar{b}) - |\bar{\alpha}||\bar{a}| \geq 0$ and the inequality is strict unless $\bar{\alpha} = \bar{\beta} = \bar{a} = \bar{b} = -1$. ☐

Remark: Because $R > 1/2$ it is always true that $-R^{-1} > -2$, but $-2 \geq -P^{-1}$ iff $1/2 \geq P$. Hence, $(-1,-1)$ is in the ZDS strip iff $1/2 \geq P$.

For a ZDS, we can usefully transform the Press–Dyson equation (2.15).

Proposition 3.2. *Assume that* X *uses plan* $\mathbf{p}$ *with* X *Press–Dyson vector* $\tilde{\mathbf{p}} = \gamma(\bar{\alpha}\mathbf{S}_X + \bar{\beta}\mathbf{S}_Y + \mathbf{1})$, $\gamma > 0$. *Let* $-Z^{-1} = \bar{\alpha} + \bar{\beta}$, *so that* $P \le Z \le R$.

For any general plan played by Y,

$$\bar{\alpha}Z(s_X - s_Y) = (s_Y - Z). \tag{3.4}$$

If $\kappa = \bar{\alpha}Z/(1 + \bar{\alpha}Z)$, *then* $1 > \kappa$ *and* κ *has the same sign as* $\bar{\alpha}$. *For any general plan played by* Y,

$$\kappa(s_X - Z) = (s_Y - Z). \tag{3.5}$$

Proof. Notice that $1 + \bar{\alpha}Z = -\bar{\beta}Z \ge Z \ge P > 0$. Multiplying Equation (2.15) by Z and substituting for $\bar{\beta}Z$ easily yields Equation (3.4) and then Equation (3.5). □

If $\bar{\alpha} = 0$, which is the equalizer case, $s_Y = Z$ and s_X is undetermined. When $\bar{\alpha} > 0$, the payoffs s_X and s_Y are on the same side of Z, while they are on opposite sides when $\bar{\alpha} < 0$. To be precise, we have the following.

Corollary 3.3. *Assume that* X *uses a plan* $\mathbf{p}$ *with* X *Press–Dyson vector* $\tilde{\mathbf{p}} = \gamma(\bar{\alpha}\mathbf{S}_X + \bar{\beta}\mathbf{S}_Y + \mathbf{1})$, $\gamma > 0$. *Let* $-Z^{-1} = \bar{\alpha} + \bar{\beta}$. *Assume that* Y *uses an arbitrary general plan.*

(a) *If* $\bar{\alpha} = 0$, *then* $s_Y = Z$. *If* $\bar{\alpha} \ne 0$, *then the following are equivalent*
  (i) $s_Y = s_X$,
  (ii) $s_Y = Z$,
  (iii) $s_X = Z$.

(b) *If* $s_Y > s_X$, *then*

$$\begin{cases} \bar{\alpha} > 0, \Rightarrow \; Z > s_Y > s_X, \\ \bar{\alpha} = 0, \Rightarrow \; Z = s_Y > s_X, \\ \bar{\alpha} < 0, \Rightarrow \; s_Y > Z > s_X. \end{cases} \tag{3.6}$$

(c) *If* $s_X > s_Y$, *then*

$$\begin{cases} \bar{\alpha} > 0, \Rightarrow \; s_X > s_Y > Z, \\ \bar{\alpha} = 0, \Rightarrow \; s_X > s_Y = Z, \\ \bar{\alpha} < 0, \Rightarrow \; s_X > Z > s_Y. \end{cases} \tag{3.7}$$

Proof. (a) If $\bar{\alpha} = 0$, then $s_Y = Z$ by Equation (3.4). If $\bar{\alpha} \ne 0$, then (i) $\Leftrightarrow$ (ii) by Equation (3.4) and (ii) $\Leftrightarrow$ (iii) by Equation (3.5).

(b), (c) If $\bar{\alpha} \ne 0$, then by Equation (3.4) $s_Y - Z$ has the same sign as that of $\bar{\alpha}(s_X - s_Y)$. □

For $Z = R$, Equation (3.5) is Equation (2.12). When $\bar{\alpha} > 0$, these are the complier strategies, i.e., the generous, good plans described in Proposition 2.6.

For $Z = P, \bar{\alpha} > 0$, the plans are firm. These were considered by Press and Dyson who called them *extortion strategies*. The name comes from the observation that whenever Y chooses a strategy so that her payoff is above P, the bonus beyond P is divided between X and Y in a ratio of $1 : \kappa$. They point out that the best reply against such an extortion play by X is for Y is to play All$C = (1, 1, 1, 1)$, which gives X a payoff above R. At first glance, it seems hard to escape from this coercive effect. I believe that the answer is for Y to play a generous good strategy like the compliers above. With repeated play, each player receives enough data to estimate statistically the strategy used by the opponent. Y's good strategy represents a credible invitation for X to switch to an agreeable plan and receive R, or else be locked below R. Hence, it undercuts the threat from X to remain extortionate.

In order to compute what happens when both players use a ZDS, we need to examine the symmetry between the two players. Let Switch : $\mathbb{R}^4 \to \mathbb{R}^4$ be defined by Switch$(x_1, x_2, x_3, x_4) = (x_1, x_3, x_2, x_4)$. Notice that Switch interchanges the vectors $\mathbf{S}_\mathrm{X}$ and $\mathbf{S}_\mathrm{Y}$. If X uses $\mathbf{p}$ and Y uses $\mathbf{q}$, then recall that the response vectors used to build the Markov matrix $\mathbf{M}$ are $\mathbf{p}$ and Switch$(\mathbf{q})$. Now suppose that the two players exchange plans so that X uses $\mathbf{q}$ and Y uses $\mathbf{p}$. Then the X response is $\mathbf{q} =$ Switch(Switch$(\mathbf{q})$) and the Y response is Switch$(\mathbf{p})$. Hence, the new Markov matrix is obtained by transposing both the second and third rows and the second and third columns. It follows that if $\mathbf{v}$ was a stationary vector for $\mathbf{M}$, then Switch$(\mathbf{v})$ is a stationary vector for the new matrix. Hence, Theorem 1.3 applied to the X Press–Dyson vector $\tilde{\mathbf{q}}$ implies that $0 = \langle \mathrm{Switch}(\mathbf{v}) \cdot \tilde{\mathbf{q}} \rangle = \langle \mathbf{v} \cdot \mathrm{Switch}(\tilde{\mathbf{q}}) \rangle$. Furthermore, if $\tilde{\mathbf{q}} = a\mathbf{S}_\mathrm{X} + b\mathbf{S}_\mathrm{Y} + g\mathbf{1} + \delta \mathbf{e}_{23}$, then Switch$(\tilde{\mathbf{q}}) = b\mathbf{S}_\mathrm{X} + a\mathbf{S}_\mathrm{Y} + g\mathbf{1} + \delta \mathbf{e}_{23}$.

For a plan $\mathbf{q}$, we define Y *Press–Dyson vector* $\tilde{\tilde{\mathbf{q}}} = \mathrm{Switch}(\tilde{\mathbf{q}}) = \mathrm{Switch}(\mathbf{q}) - \mathbf{e}_{13}$, where $\mathbf{e}_{13} = (1, 0, 1, 0)$. For any general plan for X and any limiting distribution $\mathbf{v}$ when Y uses $\mathbf{q}$ we have $\langle \mathbf{v} \cdot \tilde{\tilde{\mathbf{q}}} \rangle = 0$. The plan $\mathbf{q}$ is a ZDS associated with $(\bar{a}, \bar{b})$ in the ZDS strip when $\tilde{\tilde{\mathbf{q}}} = g(\bar{b}\mathbf{S}_\mathrm{X} + \bar{a}\mathbf{S}_\mathrm{Y} + \mathbf{1})$ with some $g > 0$.

Now we compute what happens when X and Y use ZDS plans associated, respectively, with points $(\bar{\alpha}, \bar{\beta})$ and $(\bar{a}, \bar{b})$ in the ZDS strip. This means that for some $\gamma > 0, g > 0$, $\tilde{\mathbf{p}} = \gamma(\bar{\alpha}\mathbf{S}_\mathrm{X} + \bar{\beta}\mathbf{S}_\mathrm{Y} + \mathbf{1})$ and $\tilde{\tilde{\mathbf{q}}} = g(\bar{b}\mathbf{S}_\mathrm{X} + \bar{a}\mathbf{S}_\mathrm{Y} + \mathbf{1})$. We obtain two Press–Dyson equations that hold simultaneously

$$\begin{aligned} \bar{\alpha}s_\mathrm{X} + \bar{\beta}s_\mathrm{Y} &= -1, \\ \bar{b}s_\mathrm{X} + \bar{a}s_\mathrm{Y} &= -1. \end{aligned} \tag{3.8}$$

If $\bar{\alpha} = \bar{\beta} = \bar{a} = \bar{b} = -1$, which we will call a Vertex plan $= \gamma(2(1-R), 1, 0, 1-2P)$, then the two equations are the same. Following the Remark after Lemma 3.1, a Vertex plan can occur only when $P \leq 1/2$. Clearly, $\{cd\}$ and $\{dc\}$ are both terminal sets when both players use a Vertex plan and so the payoffs depend upon the initial plays. If the two players use the same initial play as well as the same plan, then $s_\mathrm{X} = s_\mathrm{Y}$ and the single equation of (3.8) yields $s_\mathrm{X} = s_\mathrm{Y} = 1/2$.

Otherwise, Lemma 3.1 implies that the determinant $D = \bar{\beta}\bar{b} - \bar{\alpha}\bar{a}$ is positive, and by Cramer's rule, we get

$$\begin{aligned} &s_\mathrm{X} = D^{-1}(\bar{a} - \bar{\beta}), \quad s_\mathrm{Y} = D^{-1}(\bar{\alpha} - \bar{b}), \\ &\text{and so} \quad s_\mathrm{Y} - s_\mathrm{X} = D^{-1}[(\bar{\alpha} + \bar{\beta}) - (\bar{a} + \bar{b})]. \end{aligned} \tag{3.9}$$

Notice that s_X and s_Y are independent of γ and g.

Thus, when both X and Y use ZDS plans from the ZDS strip, these long-term payoffs depend only on the plans and so the results are independent of the choice of initial plays.

Proposition 3.4. *Assume that* $\tilde{\mathbf{p}} = \gamma(\bar{\alpha}\mathbf{S}_\mathrm{X} + \bar{\beta}\mathbf{S}_\mathrm{Y} + \mathbf{1})$ *and* $\tilde{\tilde{\mathbf{q}}} = g(\bar{b}\mathbf{S}_\mathrm{X} + \bar{a}\mathbf{S}_\mathrm{Y} + \mathbf{1})$*. Let* $\bar{\alpha} + \bar{\beta} = -Z_\mathrm{X}^{-1}$ *and* $\bar{a} + \bar{b} = -Z_\mathrm{Y}^{-1}$*. Assume that* $(-1, -1)$ *is not equal to both* $(\bar{\alpha}, \bar{\beta})$ *and* $(\bar{a}, \bar{b})$*.*

(a) The points $(\bar{\alpha}, \bar{\beta}), (\bar{a}, \bar{b})$ *lie on the same value line* $x + y = -Z^{-1}$*, i.e.,* $Z_\mathrm{X} = Z_\mathrm{Y}$*, iff* $s_\mathrm{X} = s_\mathrm{Y}$*. In that case,* $Z_\mathrm{X} = s_\mathrm{X} = s_\mathrm{Y} = Z_\mathrm{Y}$*.*

(b) $s_\mathrm{Y} > s_\mathrm{X}$ *iff* $Z_\mathrm{X} > Z_\mathrm{Y}$*.*

(c) Assume $Z_\mathrm{X} > Z_\mathrm{Y}$*. The following implications hold.*

$$\begin{gathered}\begin{cases}\bar{\alpha} > 0, \Rightarrow\ Z_\mathrm{X} > s_\mathrm{Y} > s_\mathrm{X}.\\ \bar{\alpha} = 0, \Rightarrow\ Z_\mathrm{X} = s_\mathrm{Y} > s_\mathrm{X}.\\ \bar{\alpha} < 0, \Rightarrow\ s_\mathrm{Y} > Z_\mathrm{X} > s_\mathrm{X}.\end{cases}\\ \begin{cases}\bar{a} > 0, \Rightarrow\ s_\mathrm{Y} > s_\mathrm{X} > Z_\mathrm{Y}.\\ \bar{a} = 0, \Rightarrow\ s_\mathrm{Y} > s_\mathrm{X} = Z_\mathrm{Y}.\\ \bar{a} < 0, \Rightarrow\ s_\mathrm{Y} > Z_\mathrm{Y} > s_\mathrm{X}.\end{cases}\end{gathered} \tag{3.10}$$

Proof. We are excluding by assumption the case when both players use Vertex plans and so we have $D > 0$.

(a) Assume $Z_\mathrm{X} = Z_\mathrm{Y}$. From Equation (3.9), we see that $s_\mathrm{Y} - s_\mathrm{X} = 0$.
When $s_\mathrm{X} = s_\mathrm{Y}$, Corollary 3.3(a) implies that $s_\mathrm{X} = s_\mathrm{Y} = Z_\mathrm{X}$. By using the XY symmetry, we see that the common value is Z_Y as well. Hence, $Z_\mathrm{X} = Z_\mathrm{Y}$ and the points lie on the same line.

(b) Since $D > 0$, (b) follows from Equation (3.9).

(c) From (b), $s_\mathrm{Y} - s_\mathrm{X} > 0$. The first part follows from Equation (3.6) with $Z = Z_\mathrm{X}$. The second follows from Equation (3.7) by using the XY symmetry with $\bar{\alpha}$, $\bar{\beta}$, Z replaced by $\bar{a}$, $\bar{b}$, Z_Y. □

Remark: If both players use a Vertex plan and the same initial play, then (a) holds with $Z_\mathrm{X} = Z_\mathrm{Y} = 1/2$.

4 Dynamics among zero-determinant strategies

In this section, we move beyond the classical question that motivated our original interest in good strategies. We consider now the evolutionary dynamics among memory one strategies. We follow Chapter 9 of Hofbauer and Sigmund [9] and Akin [2].

The dynamics that we consider takes place in the context of a symmetric two-person game, but generalizing our initial description, we merely assume that there is a set of strategies indexed by a finite set $\mathcal{I}$. When players X and Y use strategies with index $i, j \in \mathcal{I}$, respectively, the payoff to player X is given by A_{ij} and the payoff to Y by A_{ji}. Thus, the game is described by the payoff matrix $\{A_{ij}\}$. We imagine a population of players each using a particular strategy for each encounter and let π_i denote the ratio of the number of i players to the total population. The frequency vector $\{\pi_i\}$ lives in the unit simplex $\Delta \subset \mathbb{R}^{\mathcal{I}}$, i.e., the entries are non-negative and sum to 1. The vertex $v(i)$ associated with $i \in \mathcal{I}$ corresponds to a population consisting entirely of i players. We assume the population is large so that we can regard π as changing continuously in time.

Now we regard the payoff in units of *fitness*. That is, when an i player meets a j player in an interval of time dt, the payoff A_{ij} is an addition to the background reproductive rate ρ of the members of the population. So the i player is replaced by $1 + (\rho + A_{ij})dt$ i players. Averaging over the current population distribution, the expected relative reproductive rate for the subpopulation of i players is $\rho + A_{i\pi}$, where

$$\begin{aligned} A_{i\pi} &= \Sigma_{j\in\mathcal{I}}\pi_j A_{ij} \quad \text{and} \\ A_{\pi\pi} &= \Sigma_{i\in\mathcal{I}}\pi_i A_{i\pi} = \Sigma_{i,j\in\mathcal{I}}\pi_i\pi_j A_{ij}. \end{aligned} \tag{4.1}$$

The resulting dynamical system on Δ is given by the *Taylor–Jonker game dynamics equations* introduced by Taylor and Jonker [18].

$$\frac{d\pi_i}{dt} = \pi_i(A_{i\pi} - A_{\pi\pi}). \tag{4.2}$$

This system is an example of the *replicator equations* studied in great detail by Hofbauer and Sigmund [9].

We will need some general game dynamic results for later application. Fix the game matrix $\{A_{ij}\}$.

A subset A of Δ is called *invariant* if $\pi(0) \in A$ implies that the entire solution path lies in A. That is, $\pi(t) \in A$ for all $t \in \mathbb{R}$. An invariant point is is an *equilibrium*.

Each nonempty subset $\mathcal{J}$ of $\mathcal{I}$ determines the *face* $\Delta_{\mathcal{J}}$ of the simplex consisting of those $\pi \in \Delta$ such that $\pi_i = 0$ for all $i \notin \mathcal{J}$. Each face of the simplex is invariant because $\pi_i = 0$ implies that $d\pi_i/dt = 0$. In particular, for each $i \in \mathcal{I}$, the vertex $v(i)$, which represents fixation at the i strategy, is an equilibrium. In general, π is an equilibrium when, for all $i, j \in \mathcal{I}$, π_i, $\pi_j > 0$ imply $A_{i\pi} = A_{j\pi}$. This implies that $A_{i\pi} = A_{\pi\pi}$ for all i such that $\pi_i > 0$. That is, for all i in the *support* of π.

An important example of an invariant set is the *omega limit point set of an orbit*. Given an initial point $\pi \in \Delta$ with associated solution path $\pi(t)$, it is defined by intersecting the closures of the tail values.

$$\omega(\pi) = \bigcap_{t>0} \overline{\{\pi(s) : s \geq t\}}. \tag{4.3}$$

By compactness this set is nonempty. A point is in $\omega(\pi)$ iff it is the limit of some sequence $\{\pi(t_n)\}$ with $\{t_n\}$ tending to infinity. The set $\omega(\pi)$ consists of a single point π^* iff $\mathrm{Lim}_{t\to\infty}\pi(t) = \pi^*$. In that case, $\{\pi^*\}$ is an invariant point, i.e., an equilibrium.

Definition 4.1. We call a strategy i^* an *evolutionarily stable strategy* (hereafter, an ESS) when

$$A_{ji^*} < A_{i^*i^*} \quad \textit{for all } j \neq i^* \textit{ in } \mathcal{I}. \tag{4.4}$$

We call a strategy i^* an *evolutionarily unstable strategy* (hereafter, an EUS) when

$$A_{ji^*} > A_{i^*i^*} \quad \textit{for all } j \neq i^* \textit{ in } \mathcal{I}. \tag{4.5}$$

The ESS condition above is really a special case of a more general notion, see page 63 of [9], and is referred to there as a *strict Nash equilibrium*. We will not need the generalization and we use the term to avoid confusion with the strategies of Nash type considered in the previous sections.

Proposition 4.2. *If i^* is an ESS, then the vertex $v(i^*)$ is an attractor, i.e., a locally stable equilibrium, for System (4.2). In fact, there exists $\epsilon > 0$ such that*

$$1 > \pi_{i^*} \geq 1-\epsilon \implies \frac{d\pi_{i^*}}{dt} > 0. \tag{4.6}$$

Thus, near the equilibrium $v(i^*)$, which is characterized by $\pi_{i^*} = 1$, $\pi_{i^*}(t)$ increases monotonically, converging to 1 and the alternative strategies are eliminated from the population in the limit.

If i^* is an EUS, then the vertex $v(i^*)$ is a repellor, i.e., a locally unstable equilibrium, for System (4.2). In fact, there exists $\epsilon > 0$ such that

$$1 > \pi_{i^*} \geq 1-\epsilon \implies \frac{d\pi_{i^*}}{dt} < 0. \tag{4.7}$$

Thus, near the equilibrium $v(i^*)$, $\pi_{i^*}(t)$ decreases monotonically, until the system enters, and then remains in the region where $\pi_{i^*} < 1-\epsilon$.

Proof. When i^* is an ESS, $A_{i^*i^*} > A_{ji^*}$ for all $j \neq i^*$. It then follows for $\epsilon > 0$ sufficiently small that $\pi_{i^*} \geq 1-\epsilon$ implies $A_{i^*\pi} > A_{j\pi}$ for all $j \neq i^*$. If also $1 > \pi_{i^*}$, then $A_{i^*\pi} > A_{\pi\pi}$. So Equation (4.2) implies Equation (4.6).

The EUS case is similar. Notice that no solution path can cross $\Delta \cap \{\pi_{i^*} = 1-\epsilon\}$ from $\{\pi_{i^*} < 1-\epsilon\}$. □

Definition 4.3. For $\mathcal{J}$, a nonempty subset of $\mathcal{I}$, we say a strategy i *weakly dominates* a strategy j in $\mathcal{J}$ when $i, j \in \mathcal{J}$ and

$$A_{jk} \leq A_{ik} \quad \text{for all } k \in \mathcal{J}, \tag{4.8}$$

and the inequality is strict for either $k = i$ or $k = j$. If the inequalities are strict for all k, then we say that i *dominates* j in $\mathcal{J}$.

We say that $i \in \mathcal{J}$ dominates a sequence $\{j_1, \ldots, j_n\}$ in $\mathcal{J}$ when i dominates j_1 in $\mathcal{J}$ and for $p = 2, \ldots, n$, i dominates j_p in $\mathcal{J} \setminus \{j_1, \ldots, j_{p-1}\}$.

When $\mathcal{J}$ equals all of $\mathcal{I}$, we will omit the phrase "in $\mathcal{J}$."

For $i, j \in \mathcal{I}$, define the set Q_{ij} and on it the real valued function L_{ij} by

$$\begin{aligned} Q_{ij} &= \{\pi \in \Delta : \pi_i, \pi_j > 0\} \\ L_{ij}(\pi) &= \ln(\pi_i) - \ln(\pi_j). \end{aligned} \tag{4.9}$$

Lemma 4.4. (a) *If i weakly dominates j, then $dL_{ij}/dt > 0$ on the set Q_{ij}.*
(b) *If i dominates j in $\mathcal{J}$, then there exists $\epsilon > 0$ such that $dL_{ij}/dt > 0$ on the set $Q_{ij} \cap \{\pi \in \Delta : \Sigma_{k \notin \mathcal{J}} \pi_k \leq \epsilon\}$.*

Proof. Observe that

$$dL_{ij}/dt = A_{i\pi} - A_{j\pi} = \Sigma_{k \in \mathcal{I}} \pi_k (A_{ik} - A_{jk}) \tag{4.10}$$

(a) Since $\pi_i, \pi_j > 0$ in Q_{ij} and $A_{ik} - A_{jk} \geq 0$ for all k with strict inequality for $k = i$ or $k = j$, it follows that the derivative is positive.
(b) Define

$$\begin{aligned} m &= \min\{A_{ik} - A_{jk} : k \in \mathcal{J}\} > 0, \\ M &= \max\{|A_{ik} - A_{jk}| : k \notin \mathcal{J}\}, \\ \pi_{\mathcal{J}} &= \Sigma_{k \in \mathcal{J}} \pi_k, \\ \pi_{k|\mathcal{J}} &= \pi_k / \pi_{\mathcal{J}} \quad \text{for } k \in \mathcal{J}. \end{aligned} \tag{4.11}$$

Observe that $\Sigma_{k \notin \mathcal{J}} \pi_k = 1 - \pi_{\mathcal{J}}$.

For any $\pi \in Q_{ij}$

$$\begin{aligned} A_{i\pi} - A_{j\pi} &= \pi_{\mathcal{J}} \Sigma_{k \in \mathcal{J}} \pi_{k|\mathcal{J}} (A_{ik} - A_{jk}) + \Sigma_{k \notin \mathcal{J}} \pi_k (A_{ik} - A_{jk}) \\ &\geq \pi_{\mathcal{J}} m - (1 - \pi_{\mathcal{J}}) M. \end{aligned} \tag{4.12}$$

So if ϵ is chosen with $0 < \epsilon < m/(m + M)$, then $A_{i\pi} - A_{j\pi} > 0$ when $\pi \in Q_{ij} \cap \{\pi \in \Delta : (1 - \pi_{\mathcal{J}}) \leq \epsilon\}$. $\square$

Lemma 4.5. *If $\pi(t)$ is a solution path with $\pi(0) \in Q_{ij}$ and there exists $T \in \mathbb{R}$ such that $dL_{ij}/dt > 0$ on the set $Q_{ij} \cap \overline{\{\pi(t) : t \geq T\}}$, then*

$$\operatorname{Lim}_{t \to \infty} \pi_j(t) = 0. \tag{4.13}$$

Proof. By assumption, $L_{ij}(\pi(t))$ is a strictly increasing function of t for $t \geq T$. Thus, as a t tends to infinity, $L_{ij}(\pi(t))$ approaches $\ell = \sup\{L_{ij}(\pi(t)) : t \geq T\}$ with $L_{ij}(\pi(T)) < \ell \leq +\infty$.

We must prove that $\pi_j = 0$ on the omega limit set. Assume instead that $\pi^* \in \omega(\pi(0))$ with $\pi^*_j > 0$. If π^*_i were 0, then $L_{ij}(\pi(t))$ would not be bounded below on $\{\pi(t) : t \geq T\}$. Hence, π^* lies in Q_{ij} with $\ell = L_{ij}(\pi^*) < \infty$. So on the invariant set $\omega(\pi(0)) \cap Q_{ij}$, which contains π^* and so is nonempty, L_{ij} would be constantly $\ell < \infty$. Since this set is invariant, dL_{ij}/dt would equal zero. This contradicts our assumption that the derivative is positive on $\omega(\pi(0)) \cap Q_{ij}$. $\square$

Proposition 4.6. *For $i \in \mathcal{I}$, let $\pi(t)$ be a solution path with $\pi_i(0) > 0$*
(a) *If i weakly dominates j, then* $\lim_{t\to\infty} \pi_j(t) = 0$.
(b) *If i dominates the sequence* $\{j_1, \ldots, j_n\}$ *then for* $j = j_1, \ldots, j_n$, $\lim_{t\to\infty} \pi_j(t) = 0$.

Proof. (a) If $\pi_j(0) = 0$, then $\pi_j(t) = 0$ for all t and so the limit is 0. Hence, we may assume $\pi_j(0) > 0$ and so that $\pi(0) \in Q_{ij}$. By Lemma 4.4(a), $dL_{ij}/dt > 0$ on Q_{ij} and so Lemma 4.5 implies $\mathrm{Lim}_{t\to\infty}\pi_j(t) = 0$.
(b) We prove the result by induction on n.
By part (a) $\lim_{t\to\infty} \pi_j(t) = 0$, for $j = j_1$.
Now assume the limit result is true for $j = j_1, \ldots, j_{p-1}$ with $1 < p \leq n$. We prove the result for $j = j_p$.
Let $\mathcal{J} = \mathcal{I} \setminus \{j_1, \ldots, j_{p-1}\}$. By assumption, i dominates j_p in $\mathcal{J}$. Hence, with $j = j_p$, Lemma 4.4(b) implies that there exists $\epsilon > 0$ such that $dL_{ij}/dt > 0$ on the set $Q_{ij} \cap \{\pi \in \Delta : \Sigma_{k\notin\mathcal{J}}\pi_k \leq \epsilon\}$.
By induction hypothesis, there exists T such that $\Sigma_{k\notin\mathcal{J}}\pi_k(t) \leq \epsilon$ for all $t \geq T$. Hence, $\overline{\{\pi(t) : t \geq T\}} \subset \{\pi : \Sigma_{k\notin\mathcal{J}}\pi_k(t) \leq \epsilon\}$.
As in part (a), we can assume $\pi \in Q_{ij}$ and then apply Lemma 4.5 to conclude $\lim_{t\to\infty} \pi_j(t) = 0$. This completes the inductive step. □

Now we specialize to the iterated Prisoner's Dilemma. By a *strategy*, we will mean a plan **p** together with an initial play, pure or mixed. Recall that a good (or agreeable) strategy is a good (respectively agreeable) plan together with initial cooperation.

To apply the Taylor–Jonker dynamics to our case, we suppose that $\mathcal{I}$ indexes a finite collection of strategies. We then use

$$A_{ij} = s_X \text{ so that } A_{ji} = s_Y. \tag{4.14}$$

That is, when the X player uses the i strategy and the Y player uses the j strategy, the players receive the payoffs s_X and s_Y, respectively, as additions to their reproductive rate. When the associated Markov matrix is convergent, there is a unique terminal set, and the long-term payoffs, s_X, s_Y depend only on the plans and not on the initial plays.

Theorem 4.7. *Let $\mathcal{I}$ index a finite set of strategies for the iterated Prisoner's Dilemma. Suppose that associated with $i^* \in \mathcal{I}$ is a good strategy $\mathbf{p}^{i*}$. If for no other $j \in \mathcal{I}$ is the plan $\mathbf{p}^j$ agreeable, then i^* is an ESS for the associated game $\{A_{ij} : i, j \in \mathcal{I}\}$ and so the vertex $v(i^*)$ is an attractor for the dynamic.*

Proof. Since i^* is associated with an agreeable strategy, $A_{i^*i^*} = R$. Since $\mathbf{p}^{i^*}$ is good and $\mathbf{p}^j$ is not agreeable for $j \neq i^*$, it follows from Corollary 2.2(a) that $A_{ji^*} < R$ for $j \neq i^*$. Thus, i^* is an ESS. □

There are other cases of ESS that are far from good.

Lemma 4.8. (a) *Assume that* X *uses a plan* $\mathbf{p} = (p_1, p_2, 0, 0)$ *with* $p_1, p_2 < 1$. *If* Y *uses any plan* $\mathbf{q}$ *that is not firm, then*

$$s_Y < P < s_X. \tag{4.15}$$

(b) *Assume that* X *uses a plan* $\mathbf{p} = (1, p_2, 0, 0)$ *with* $p_2 < 1$. *If* Y *uses any plan* $\mathbf{q}$ *that is neither firm nor agreeable, then Equation (4.15) holds.*

(c) *Assume* $P < 1/2$ *and that* X *uses a firm, non-exceptional ZDS with* $\bar{\alpha} < 0$. *If* Y *uses any plan* $\mathbf{q}$ *that is not firm, then Equation (4.15) holds.*

Proof. (a) and (b) Since $p_3 = p_4 = 0$, the set $\{dc, dd\}$ is closed. If $q_4 > 0$, then $\{dd\}$ is not closed and so is not a terminal set.

(a) Since $p_2 < 1$, there is an edge from cd to either dc or dd. Hence, cd is transient. Similarly, $p_1 < 1$ implies that cc is transient. Hence, for any stationary distribution $\mathbf{v}$, $v_1 = v_2 = 0$. Since $\mathbf{q}$ is not firm, $q_4 > 0$ and so $v_4 < 1$. Hence, $s_Y = v_4 P < P$ and $s_X = v_3 + v_4 P = (1 - v_4) + v_4 P > P$.

(b) As before, $p_2 < 1$ implies that cd is transient. Now $\mathbf{q}$ is not agreeable and so $q_1 < 1$. This implies that there is an edge from cc to the transient state cd and so cc is transient. The proof is completed as in (a).

(c) Because $P < 1/2$, the smallest entry in $1/2(\mathbf{S}_X + \mathbf{S}_Y)$ is P and so $1/2(s_X + s_Y) \leq P$ can only happen when $v_4 = 1$, which implies that $s_X = s_Y = P$. This requires that Y play a firm plan so that $\{dd\}$ is a terminal set. Compare Proposition 1.1.

From Equation (3.5) we see that with $\bar{\alpha} + \bar{\beta} = -Z^{-1}$ and $\kappa = \bar{\alpha}Z/(1 + \bar{\alpha}Z)$

$$\frac{1}{2}(1 + \kappa)(s_X - Z) = \left(\frac{1}{2}(s_X + s_Y) - Z\right). \tag{4.16}$$

When $Z = P$, $P < 1/2$ and $-1 \geq \bar{\alpha}$ imply that $(1 + \kappa) = (1 + 2\bar{\alpha}P)/(1 + \bar{\alpha}P) > 0$. Hence, $s_X \leq P$ implies that $1/2(s_X + s_Y) \leq P$. Since the Y plan is not firm, this does not happen. Hence, $s_X > P$. Since $\kappa < 0$, Equation (3.5) implies that $s_Y < P$. ☐

Theorem 4.9. *Let* $\mathcal{I}$ *index a finite set of strategies for the iterated Prisoner's Dilemma.*

(a) *Suppose that associated with* $i^* \in \mathcal{I}$ *is a plan* $\mathbf{p}^{i*} = (p_1, p_2, 0, 0)$ *with* $p_1, p_2 < 1$ *together with any initial play. If for no other* $j \in \mathcal{I}$ *is the plan* $\mathbf{p}^j$ *firm, then* i^* *is an ESS for the associated game* $\{A_{ij} : i, j \in \mathcal{I}\}$.

(b) *Suppose that associated with* $i^* \in \mathcal{I}$ *is a plan* $\mathbf{p}^{i*} = (1, p_2, 0, 0)$ *with* $p_2 < 1$ *together with any initial play. If for no other* $j \in \mathcal{I}$ *is the plan* $\mathbf{p}^j$ *either agreeable or firm, then* i^* *is an ESS for the associated game* $\{A_{ij} : i, j \in \mathcal{I}\}$.

(c) *Assume that* $P < 1/2$. *Suppose that associated with* $i^* \in \mathcal{I}$ *is a firm, non-exceptional ZDS with* $\bar{\alpha} < 0$ *together with any initial play. If for no other* $j \in \mathcal{I}$ *is the plan* $\mathbf{p}^j$ *firm, then* i^* *is an ESS for the associated game* $\{A_{ij} : i, j \in \mathcal{I}\}$.

Proof. (a) If both players use p^{i^*}, then there is an edge from dc to dd and so dc, cd, and cc are all transient. Thus, $\{dd\}$ is the unique terminal set and so

$A_{i^*i^*} = P$ regardless of the initial plays. By Lemma 4.8(a), $A_{ji^*} < P$ for all $j \neq i^*$.

(b) If both players use p^{i^*}, then there are edges from cd to dd and from dc to dd. The two terminal sets are $\{cc\}$ and $\{dd\}$. Hence, $R \geq A_{i^*i^*} \geq P$. This time Lemma 4.8(b) implies that $A_{ji^*} < P$ for any $j \neq i^*$.

(c) $A_{i^*i^*} = Z_{i^*} = P$ since the i^* plan is firm. Lemma 4.8(c) implies that $A_{ji^*} < P$ for any $j \neq i^*$. ☐

Thus, $p^{i^*} = \text{All}D = (0,0,0,0)$ with any initial play is an ESS when played against plans that are not firm. If $p^{i^*} = \text{Grim} = (1,0,0,0)$, then with any initial play i^* is an ESS when played against strategies that are neither agreeable nor firm.

At the other extreme, we have the following.

Theorem 4.10. *Let $\mathcal{I}$ index a finite set of strategies for the iterated Prisoner's Dilemma. Assume that $P < 1/2$. Suppose that associated with $i^* \in \mathcal{I}$ is an extortionate plan $\mathbf{p}^{i*}$ together with initial defection. That is, $\mathbf{p}^{i^*}$ is a firm ZDS with $\bar{\alpha} > 0$. If for no other $j \in \mathcal{I}$ is the plan $\mathbf{p}^j$ firm, then i^* is an EUS for the associated game $\{A_{ij} : i, j \in \mathcal{I}\}$ and so the vertex $v(i^*)$ is a repellor for the dynamic.*

Proof. Because $P < 1/2$, the smallest entry in $1/2(\mathbf{S}_\mathrm{X} + \mathbf{S}_\mathrm{Y})$ is P and so $s_\mathrm{X}, s_\mathrm{Y} \leq P$ implies $s_\mathrm{X} = s_\mathrm{Y} = P$ and this can only happen when $v_4 = 1$, which requires that Y play a firm plan so that $\{dd\}$ is a terminal set. Compare Proposition 1.1.

Since the i^* strategy is firm with initial defection, $A_{i^*i^*} = P$.

If Y uses any plan that is not firm then Equation (3.5) with $z = P$ and $\bar{\alpha} > 0$ shows that if $s_\mathrm{Y} \leq P$ then $s_\mathrm{X} \leq P$ as well. Because $P < 1/2$ this can only happen when $s_\mathrm{X} = s_\mathrm{Y} = P$ and $v_4 = 1$. But the Y plan is not firm. It follows that $s_\mathrm{Y} > P$. Thus, for any $j \neq i^*$, $A_{ji^*} > A_{i^*i^*}$. This says that strategy i^* is an EUS. ☐

We now specialize to the case when all the strategies indexed by $\mathcal{I}$ are ZDS s with the exceptional strategies excluded. We can thus regard $\mathcal{I}$ as listing a finite set of points $(\bar{\alpha}_i, \bar{\beta}_i)$ in the ZDS strip and, except for the Vertex plans, we may disregard the initial plays. We define $Z_i = -(\bar{\alpha}_i + \bar{\beta}_i)^{-1}$. That is, the point $(\bar{\alpha}_i, \bar{\beta}_i)$ lies on the value line $x + y = -(Z_i)^{-1}$.

X uses $\mathbf{p}$ associated with $(\bar{\alpha}_i, \bar{\beta}_i)$ when $\tilde{\mathbf{p}} = \gamma_i(\bar{\alpha}_i\mathbf{S}_\mathrm{X} + \bar{\beta}_i\mathbf{S}_\mathrm{Y} + \mathbf{1})$ and Y uses $\mathbf{q}$ associated with $(\bar{\alpha}_j, \bar{\beta}_j)$ when $\tilde{\tilde{\mathbf{q}}} = \gamma_j(\bar{\beta}_j\mathbf{S}_\mathrm{X} + \bar{\alpha}_j\mathbf{S}_\mathrm{Y} + \mathbf{1})$ for some $\gamma_i, \gamma_j > 0$. Notice the XY switch.

If both players use a Vertex plans with the same initial plays, then $(\bar{\alpha}_i, \bar{\beta}_i) = (-1, -1)$ and $A_{ii} = 1/2 = Z_i$. Recall that $(-1, -1)$ lies in the ZDS strip iff $P \leq 1/2$.

Otherwise, we apply Equation (3.8) with $(\bar{\alpha}, \bar{\beta}) = (\bar{\alpha}_i, \bar{\beta}_i)$ and $(\bar{a}, \bar{b}) = (\bar{\alpha}_j, \bar{\beta}_j)$. Then from Equation (3.9) we get, for $i \neq j$

$$\begin{aligned} A_{ij} &= s_\mathrm{X} = K_{ij}(\bar{\alpha}_j - \bar{\beta}_i) \\ \text{with} \quad K_{ij} &= K_{ji} = (\bar{\beta}_i\bar{\beta}_j - \bar{\alpha}_i\bar{\alpha}_j)^{-1} > 0. \end{aligned} \tag{4.17}$$

Note that these payoffs are independent of the choice of γ_i, γ_j as well as the initial plays.

By Proposition 3.4(a)

$$A_{ii} = Z_i \quad \text{for all } i \in \mathcal{I}. \tag{4.18}$$

We begin with some degenerate cases. For convenience, we exclude the Vertex plans.

First, if all of the points $(\bar{\alpha}_i, \bar{\beta}_i)$ lie on the same value line $x + y = -Z^{-1}$, i.e., all the Z_is are equal, then by Proposition 3.4(a) $A_{ij} = Z$ for all i, j and so $d\pi/dt = 0$ and every population distribution is an equilibrium. In general, if for two strategies i, j $A_{ij} = A_{ji} = Z$, then by Proposition 3.4(a) both points lie on $x + y = -Z^{-1}$ and it follows that $A_{ii} = A_{jj} = Z$ as well. In general, if $\mathcal{I}_Z = \{i : Z_i = Z\}$ contains more than one $i \in I$, then the dynamics is degenerate on the face $\Delta_{\mathcal{I}_Z}$ of the simplex.

Second, if all of the points satisfy $\bar{\alpha}_i = 0$, then all the strategies are equalizer strategies. In this case, the payoff matrix need not be constant but A_{ij} depends only on j. This implies that for all i $A_{i\pi} = A_{\pi\pi}$ and so again $d\pi/dt = 0$ and every population distribution is an equilibrium.

We will now see that the line $\bar{\alpha} = 0$ separates different interesting dynamic behaviors.

Theorem 4.11. *Let $\mathcal{I}$ index a set of non-exceptional ZDS plans. Thus, each $i \in \mathcal{I}$ is associated with a point $(\bar{\alpha}_i, \bar{\beta}_i)$ in the ZDS strip and $\bar{\alpha}_i + \bar{\beta}_i = -(Z_i)^{-1}$.*

Assume either

Case (+): $\bar{\alpha}_i > 0$ for all $i \in \mathcal{I}$ and for some $i^* \in \mathcal{I}$, $Z_{i^*} > Z_j$ for all $j \neq i^*$;

or

Case (−): $\bar{\alpha}_i < 0$ for all $i \in \mathcal{I}$ and for some $i^* \in \mathcal{I}$, $Z_{i^*} < Z_j$ for all $j \neq i^*$.

The strategy i^* is an ESS, and if $\pi_{i^*}(0) > 0$, then the solution path converges to the vertex $v(i^*)$.

Proof. List the strategies $j_1, \ldots, j_n$ of $\mathcal{I} \setminus \{i^*\}$ so that in Case(+) $Z_{j_1} \leq Z_{j_2} \leq \cdots \leq Z_{j_n} < Z_{i^*}$ and in Case(−) $Z_{j_1} \geq Z_{j_2} \geq \cdots \geq Z_{j_n} > Z_{i^*}$. For both cases, we apply Proposition 3.4. It first implies that if $Z_i = Z_j$ then

$$A_{ii} = Z_i = A_{ji} = A_{ij} = Z_j = A_{jj}. \tag{4.19}$$

Case(+): If $Z_i > Z_j$, then, because $\bar{\alpha}_i, \bar{\alpha}_j > 0$, Proposition 3.4 implies that

$$A_{ii} = Z_i > A_{ji} > A_{ij} > Z_j = A_{jj}. \tag{4.20}$$

Hence, if $Z_i > Z_k \geq Z_j$, then $A_{ii} > A_{ji}$ and $A_{ik} > A_{kk} \geq A_{jk}$.

It follows that i^* dominates the sequence $\{j_1, \ldots, j_n\}$. Hence, Proposition 4.6(b) implies that $\mathrm{Lim}_{t\to\infty}\pi_j(t) = 0$ for $j = j_1, \ldots, j_n$ when $\pi_{i^*}(0) > 0$. Consequently, $\pi_{i^*}(t) = 1 - \Sigma_{p=1}^{n}\pi_{j_p}(t)$ tends to 1. That is, $\pi(t)$ converges to $v(i^*)$.

Case(−): If $Z_i < Z_j$, then, because $\bar{\alpha}_i, \bar{\alpha}_j < 0$, Proposition 3.4 implies that

$$A_{ij} > Z_j = A_{jj} > A_{ii} = Z_i > A_{ji}. \tag{4.21}$$

It again follows that i^* dominates the sequence $\{j_1, \ldots, j_n\}$ and convergence to $v(i^*)$ again follows from Proposition 3.4.

In both cases, it is clear that i^* is an ESS. □

Thus, when only $\bar{\alpha} > 0$ ZDS plans are competing with one another, the ones on the highest value line win. Among $\bar{\alpha} < 0$ ZDS plans, the ones on the lowest value line win.

The local stability of an ESS good strategy will not be global when both signs occur. To illustrate this, consider the case of two strategies indexed by $\mathcal{I} = \{1, 2\}$. Letting $w = \pi_1$, it is an easy exercise to show that Equation (4.2) reduces to

$$\frac{dw}{dt} = w(1-w)[(A_{11} - A_{21})w + (A_{12} - A_{22})(1-w)]. \tag{4.22}$$

Proposition 4.12. *Assume that $Z_1 > Z_2$ and that $\bar{\alpha}_1 \cdot \bar{\alpha}_2 < 0$. There is an equilibrium population $\pi^* = (w^*, (1-w^*))$ that contains both strategies with*

$$w^*/(1-w^*) = (A_{22} - A_{12})/(A_{11} - A_{21}). \tag{4.23}$$

This equilibrium is stable if $\bar{\alpha}_1 < 0$ and is unstable if $\bar{\alpha}_1 > 0$.

Proof. If $\bar{\alpha}_1 < 0$ and $\bar{\alpha}_2 > 0$, then Proposition 3.4 implies that $A_{11} - A_{21} = Z_1 - A_{21} < 0$ and $A_{12} - A_{22} = A_{12} - Z_2 > 0$. Reversing the signs reverses the inequalities. The result then easily follows from Equation (4.22). Just graph the linear function of w in the brackets and observe where the result is positive or negative. □

Question 4.13. Suppose we restrict to the case where $\mathcal{I}$ indexes ZDS's lying on different value lines to avoid degeneracies. We ask:

– How large a population can coexist? If N is the size of $\mathcal{I}$, the number of competing strategies, then for what N do there exist examples with an interior equilibrium, that is, an equilibrium π such that $\pi_i > 0$ for all $i \in \mathcal{I}$? When is there a locally stable interior equilibrium? For how large an N can *permanence* occur (see Section 3 of [9]), that is, the boundary of Δ be a repellor? The *Brouwer fixed point theorem* implies that such a permanent system always admits an interior equilibrium. When an interior equilibrium does not exist, there is always some sort of dominance among the mixed strategies of the game $\{A_{ij}\}$. See [1] and [3].
– Can there exist a stable, closed invariant set containing no equilibria, e.g., a stable limit cycle?

There is alternative version of the dynamics that explicitly considers for X not the payoff s_X but the advantage that X has over Y. That is, the addition to the growth rate is given not by s_X but by the difference $s_X - s_Y$. This amounts to replacing A_{ij} by the anti-symmetric matrix $S_{ij} = A_{ij} - A_{ji}$ so that the game becomes zero-sum. In this case, we define $\xi_i = -Z_i^{-1} = \bar{\alpha}_i + \bar{\beta}_i$. Thus, ξ_i varies in the interval $[-P^{-1}, -R^{-1}]$. Define

$\xi_\pi = \Sigma_{i\in\mathcal{I}}\pi_i\xi_i$. From Equation (4.17), we get

$$S_{ij} = K_{ij}(\xi_j - \xi_i), \tag{4.24}$$

where we let $K_{ii} = 1$ for all $i \in \mathcal{I}$.

Since $\{S_{ij}\}$ is antisymmetric, $S_{\pi\pi} = 0$.

For this system, the behavior is always like the $\bar{\alpha} < 0$ case for the previous system.

Theorem 4.14. *Let $\mathcal{I}$ index a finite list of non-exceptional ZDS strategies, with at most one using a Vertex plan. For the system with*

$$\frac{d\pi_i}{dt} = \pi_i(S_{i\pi} - S_{\pi\pi}) = \pi_i S_{i\pi}, \tag{4.25}$$

we have

$$\begin{aligned} &\frac{d\xi_\pi}{dt} \leq 0, \\ &\textit{with equality iff} \quad \pi_i, \pi_j > 0 \Longrightarrow \xi_i = \xi_j. \end{aligned} \tag{4.26}$$

Assume now that $Z_{i^*} < Z_j$, or equivalently $\xi_{i^*} < \xi_j$ for all $j \neq i^*$. The strategy i^* is an ESS, and if $\pi_{i^*}(0) > 0$, then the solution path converges to the vertex $v(i^*)$.

Proof. Because K_{ij} is symmetric and positive, $d\xi_\pi/dt$ equals

$$\begin{aligned} &- \Sigma_{i,j\in\mathcal{I}}\pi_i\pi_j K_{ij}\xi_i(\xi_i - \xi_j) \\ &= -\frac{1}{2}\left[\Sigma_{i,j\in\mathcal{I}}\pi_i\pi_j K_{ij}\xi_i(\xi_i - \xi_j) - \Sigma_{i,j\in\mathcal{I}}\pi_j\pi_j K_{ij}\xi_j(\xi_i - \xi_j)\right] \\ &= -\frac{1}{2}\Sigma_{i,j\in\mathcal{I}}\pi_i\pi_j K_{ij}(\xi_i - \xi_j)^2 \leq 0. \end{aligned} \tag{4.27}$$

Equality holds iff $\pi_i\pi_j(\xi_i - \xi_j)^2 = 0$ for all $i, j \in \mathcal{I}$. That is, when $\xi_i = \xi_j$ for all i, j with $\pi_i, \pi_j > 0$.

If $\xi_i < \xi_j$, then

$$S_{ij} > 0 = S_{jj} = S_{ii} > S_{ji}. \tag{4.28}$$

If $\xi_i < \xi_k \leq \xi_j$, then $S_{ik} > S_{kk} \geq S_{jk}$. Let $\{j_1, \ldots, j_n\}$ list $\mathcal{I} \setminus \{i^*\}$ with $\xi_{j_1} \geq \cdots \geq \xi_{j_n}$. As in Case(−) of Theorem 4.11, it follows that i^* dominates the sequence $\{j_1, \ldots, j_n\}$. If $\pi_{i^*}(0) > 0$, then $\pi(t)$ converges to $v(i^*)$ by Proposition 3.4. □

References

[1] E. Akin, Domination or equilibrium, *Math. Biosciences* **50** (1980), 239–250.

[2] E. Akin, The differential geometry of population genetics and evolutionary games, in: S. Lessard (ed.), *Mathematical and Statistical Developments of Evolutionary Theory*, pp. 1–93, Kluwer, Dordrecht, 1990.

[3] E. Akin and J. Hofbauer, Recurrence of the unfit, *Math. Biosciences* **61** (1982), 51–63.

[4] R. Axelrod, *The Evolution of Cooperation*, Basic Books, New York, NY, 1984.

[5] M. Boerlijst, M. Nowak and K. Sigmund, Equal pay for all prisoners, *Amer. Math. Monthly* **104** (1997), 303–305.

[6] M. D. Davis, *Game Theory: A Nontechnical Introduction*, Dover Publications, Mineola, NY, 1983.

[7] G. Hardin, The tragedy of the commons, *Science* **162** (1968), 1243–1248.

[8] C. Hilbe, M. Nowak and K. Sigmund, The evolution of extortion in iterated Prisoner's Dilemma games, *PNAS* **110** (2013), no. 17, 6913–6918.

[9] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*, Cambridge University Press, Cambridge, UK, 1998.

[10] S. Karlin and H. Taylor, *A First Course in Stochastic Processes*, 2nd edition, Academic Press, New York, NY, 1975.

[11] J. Maynard Smith, *Evolution and the Theory of Games*, Cambridge University Press, Cambridge, UK, 1982.

[12] M. Nowak, *Evolutionary Dynamics*, Harvard University Press, Cambridge, MA, 2006.

[13] K. Sigmund, *Games of life*, Oxford University Press, Oxford, UK, 1993.

[14] K. Sigmund, *The Calculus of Selfishness*, Princeton University Press, Princeton, NJ, 2010.

[15] W. Press and F. Dyson, Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent, *PNAS* **109** (2012), no. 26, 10409–10413.

[16] A. Stewart and J. Plotkin, Extortion and cooperation in the Prisoner's Dilemma, *PNAS* **109** (2012), no. 26, 10134–10135.

[17] P. D. Straffin, *Game Theory and Strategy*, Mathematical Association of America, Washington, DC, 1993.

[18] P. Taylor and L. Jonker, Evolutionarily stable strategies and game dynamics, *Math. Biosciences* **40** (1978), 145–156.

[19] J. Von Neumann and O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton University Press, Princeton, NJ, 1944.

Marie-Claude Arnaud*‡

Lyapunov exponents for conservative twisting dynamics: a survey

Abstract: Finding special orbits (as periodic orbits) of dynamical systems by variational methods, and especially by minimization methods, is an old method (just think of the geodesic flow). More recently, new results concerning the existence of minimizing sets and minimizing measures were proved in the setting of conservative twisting dynamics. These twisting dynamics include geodesic flows as well as the dynamics close to a completely elliptic periodic point of a symplectic diffeomorphism where the torsion is positive definite (this implies the existence of a normal form $(\theta, r) \mapsto (\theta + \beta r + o(r), r + o(r))$ with β positive definite). The two aspects of this theory are called the Aubry–Mather theory and the weak KAM theory. They were built by Aubry & Mather in the 1980s in the two-dimensional case and by Mather, Mañé, and Fathi in the 1990s in higher dimension.

We will explain what are the conservative twisting dynamics and summarize the existence results of minimizing measures. Then we will explain more recent results concerning the link between different notions for minimizing measures for twisting dynamics:
- their Lyapunov exponents,
- their Oseledets splitting, and
- the shape of the support of the measure.

The main question in which we are interested is: given some minimizing measure of a conservative twisting dynamics, is there a link between the geometric shape of its support and its Lyapunov exponents? Or can we deduce the Lyapunov exponents of the measure from the "shape" of the support of this measure?

Some proofs, but not all of them, will be provided. Some questions are raised in the last section.

Keywords: twist maps, Hamiltonian dynamics, Tonelli Hamiltonians, Lagrangian functions, Lyapunov exponents, minimizing orbits and measures, Green bundles, weak KAM theory, contingent and paratangent cones.

*ANR-12-BLAN-WKBHJ

‡membre de l'Institut universitaire de France

1 Twisting conservative dynamics

All the dynamics we study here are defined on the cotangent bundle T^*M of some closed manifold M, endowed with its usual symplectic form ω. More precisely, if $q = (q_1, \dots, q_n)$ are some coordinates on M, we complete them with their dual coordinates $p = (p_1, \dots, p_n)$ to obtain some coordinates on T^*M: if $\lambda \in T^*M$ is a 1-form on M, then its coordinates $p_1, \dots, p_n$ are given by $\lambda = \sum_{i=1}^n p_i dq_i$. The expression of the symplectic form in these coordinates is $\omega = dq \wedge dp = \sum_{i=1}^n dq_i \wedge dp_i$. A change of coordinates of M doesn't change the symplectic form ω and then the definition is correct. We will generally use the notation (q, p) for such coordinates.

When $M = \mathbb{T}^n$, we will identify T^*M with the $2n$-dimensional annulus $\mathbb{A}_n = \mathbb{T}^n \times \mathbb{R}^n$.

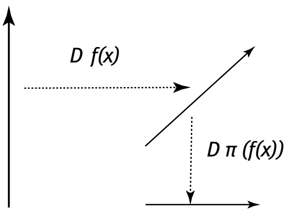
Let us recall that a diffeomorphism f of T^*M is *symplectic* if it preserves the symplectic form: $f^*\omega = \omega$.

1.1 A local notion: the twist condition

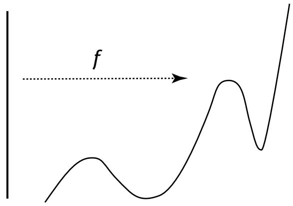
Notation 1.1. We denote by $\pi : T^*M \to M$ the canonical projection $(q, p) \mapsto q$.

At every $x = (q, p) \in T^*M$, we define the vertical subspace $V(x) = \ker D\pi(x) \subset T_x(T^*M)$ as being the tangent subspace at x to the fiber T_q^*M.

Example 1.2 A symplectic C^1 diffeomorphism $f : \mathbb{A}_1 \to \mathbb{A}_1$ of the two-dimensional annulus *is a positive symplectic twist map* if
- it is homotopic to identity;
- it twists the vertical to the positive side: $\forall x \in \mathbb{A}_1, D(\pi \circ f)(x).\binom{0}{1} > 0$.



There exists of course a notion of *negative symplectic twist map*. The notion of twist map that we introduce is local, but the (local) twist condition implies a global property: when we unfold the cylinder (i.e., we are in the universal covering $\mathbb{R}^2$ of $\mathbb{A}_1$ and we consider a lift of the twist map), the image of a fiber $\{q\} \times \mathbb{R}$ by the lift of a symplectic twist map is then a graph above a part of of $\mathbb{R} \times \{0\}$.



For example, the map $f_0 : (q,p) \mapsto (q+p,p)$ is a symplectic twist map of $\mathbb{A}_1$.

When thinking of a possible extension of the notion of twisting dynamics to higher dimension, the first possibility is to ask that the image of any vertical $V(x)$ by the tangent dynamics Df is transverse to the vertical $V(f(x))$. If we express Df in two charts of coordinates (q,p),

$$Df = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

this is equivalent to ask that $\forall x, \det b(x) \neq 0$. When $M = \mathbb{T}^n$, such diffeomorphims were introduced and studied by M. Herman in [16], where he called them *monotone*. When $\dim M \geq 2$, the monotonicity condition doesn't imply that the image of a fiber is a graph above the zero section, even if $M = \mathbb{T}^n$ and if we unfold the $2n$-dimensional annulus. For example, the map $f : \mathbb{A}_2 \to \mathbb{A}_2$ defined for $(q,p) \in \mathbb{T}^2 \times \mathbb{R}^2 = \mathbb{C}/(\mathbb{Z} + i\mathbb{Z}) \times \mathbb{R}^2$ by $f(q,p) = (q + e^{p_1 - ip_2}, p)$ is monotone but the projection of the restriction of its lift to any fiber is not injective.

If f is a twist map of the two-dimensional annulus, f^2 is not necessarily a twist map: the twist condition is just valuable for "small times" (here time 1).

Hence for Hamiltonians, we will translate the twist condition for small times. In coordinates (q,p), the Hamilton's equations for $H \in C^2(T^*M, \mathbb{R})$ are:

$$\dot{q} = \frac{\partial H}{\partial p}(q,p); \quad \dot{p} = -\frac{\partial H}{\partial q}(q,p).$$

Let us denote the Hamiltonian flow of H by (φ_t^H) and let $(\delta q, \delta p)$ be an infinitesimal solution, i.e., $\begin{pmatrix} \delta q(t) \\ \delta p(t) \end{pmatrix} = D\varphi_t(q(0), p(0)) \cdot \begin{pmatrix} \delta q(0) \\ \delta p(0) \end{pmatrix}$. By differentiating the Hamilton's equations, we obtain $\delta \dot{q} = \frac{\partial^2 H}{\partial q \partial p}\delta q + \frac{\partial^2 H}{\partial p^2}\delta p$ and then

$$D(\pi \circ \varphi_t)(q(0), p(0)) \cdot \begin{pmatrix} 0 \\ \delta p \end{pmatrix} = t\frac{\partial^2 H}{\partial p^2}(q(0), p(0))\delta p + o(t).$$

We will say that the Hamiltonian H satisfies the twist condition if at every point $\frac{\partial^2 H}{\partial p^2}$ is non degenerate. In this case, even for small times, the Hamiltonian flow is not necessarily a twist map; indeed, the $o(t)$ above is not uniform in (q,p).

1.2 Global notions: globally positive diffeomorphisms and Tonelli Hamiltonians

Unfortunately, we are able to do nothing with the local definition of twisting dynamics that we gave in subsection 1.1.

There are two problems:

(1) we need to find some special invariant subsets for the dynamics;
(2) we want to say something about the Lyapunov exponents along these invariant subsets.

In general, there are two main ways to find invariant subsets for those dynamics: perturbative methods and variational methods. Perturbative methods, as KAM theorems are, are valuable close to completely integrable dynamics (see [16] for the definition in the case of the $2n$-dimensional annulus). C. Golé gives in [14], section 27.B, a similar condition, which he calls "asymptotic linearity," that makes possible the use of variational methods in this perturbative case. But we will not explain the perturbative case in this survey. We will only work in the so-called *coercive* case (see section 27.B of [14] for example).

More precisely, we will make some assumptions such that

- we can associate a function $\mathcal{F}$ to the dynamics, such that the critical points of $\mathcal{F}$ are in some sense the orbits for the dynamics;
- the function $\mathcal{F}$ admits some minima, and then some "minimizing orbits." For the symplectic twist maps of the two-dimensional annulus, these minimizing orbits are the heart of the theory that S. Aubry and J. Mather independently developed at the beginning of the 1980s (see [6] and [22]).

That is why we introduce the following definitions. The first one comes from [19] and [14], and the second one is very classical.

1.2.1 Globally positive diffeomorphisms

Definition 1.3. A *globally positive* diffeomorphism of $\mathbb{A}_n$ is a symplectic C^1 diffeomorphism $f : \mathbb{A}_n \to \mathbb{A}_n$ that is homotopic to $\mathrm{Id}_{\mathbb{A}_n}$ and that has a lift $F : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \times \mathbb{R}^n$ that admits a C^2 generating function $S : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ such that

- $\forall k \in \mathbb{Z}^n, S(q + k, Q + k) = S(q, Q)$;
- there exists $\alpha > 0$ such that $\frac{\partial^2 S}{\partial q \partial Q}(q, Q)(v, v) \le -\alpha \|v\|^2$;
- F is implicitly given by

$$F(q,p) = (Q,P) \iff \begin{cases} p = -\dfrac{\partial S}{\partial q}(q,Q) \\ P = \dfrac{\partial S}{\partial Q}(q,Q) \end{cases}.$$

Example 1.4. The diffeomorphism $F_0 : (q,p) \in \mathbb{R}^n \times \mathbb{R}^n \mapsto (q+p,p) \in \mathbb{R}^n \times \mathbb{R}^n$ is the lift of a globally positive diffeomorphism f_0 of $\mathbb{A}_n$, and a generating function associated to F_0 is defined by $S_0(q,Q) = \frac{1}{2}\|q-Q\|^2$.

If f, F satisfy the above hypotheses, the restriction to any fiber $\{q\} \times \mathbb{R}^n$ of $\pi \circ F$ and $\pi \circ F^{-1}$ are diffeomorphisms (a proof is given in [14]). In particular, this implies that f is (locally) monotone.

Moreover, for every $k \geq 2$, $q_0, q_k \in \mathbb{R}^n$, the function $\mathcal{F} : (\mathbb{R}^n)^{k-1} \to \mathbb{R}$ defined by $\mathcal{F}(q_1, \ldots, q_{k-1}) = \sum_{j=1}^{k} S(q_{j-1}, q_j)$ has a minimum, and at every critical point for $\mathcal{F}$, the following sequence is a piece of orbit for F:

$$\left(q_0, -\frac{\partial S}{\partial q}(q_0,q_1)\right), \left(q_1, \frac{\partial S}{\partial Q}(q_0,q_1)\right),$$
$$\left(q_2, \frac{\partial S}{\partial Q}(q_1,q_2)\right), \ldots, \left(q_k, \frac{\partial S}{\partial Q}(q_{k-1},q_k)\right).$$

For the map F_0 defined in Example 1.4, the function $\mathcal{F}_0$ defined by

$\mathcal{F}_0(q_1, \ldots, q_{k-1}) = \frac{1}{2}\sum_{j=1}^{k} \|q_{j-1} - q_j\|^2$ attains its minimum at its unique critical point $(q_1, \ldots, q_{k-1}) = (q_0 + \frac{q_k - q_0}{k}, q_0 + 2\frac{q_k-q_0}{k}, \ldots, q_0 + (k-1)\frac{q_k-q_0}{k})$ and the corresponding piece of orbit is:

$$\left(q_0, \frac{q_k - q_0}{k}\right), \left(q_1, \frac{q_k - q_0}{k}\right), \ldots, \left(q_k, \frac{q_k - q_0}{k}\right).$$

Let us discuss a little on the condition on $\frac{\partial^2 S}{\partial q \partial Q}$. If the matrix of Df in coordinates (q,p) is $Df = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, we have

$$\big(b(q,p)\big)^{-1} = -\frac{\partial^2 S}{\partial q \partial Q}(q,Q).$$

Hence, the condition that we gave for the partial derivatives of S can be rewritten in terms of matrices: $b^{-1} + {}^t b^{-1} \geq \alpha \mathbf{1}$, where $\mathbf{1}$ is the identity matrix and we use the usual order for the symmetric matrices.

The reader could think of some other possible notions of global twist, for which ${}^t b^{-1} + b^{-1}$ is indefinite. But in this case, very pathological phenomena can occur; M. Herman showed very strange phenomena in the case of a “normal indefinite torsion” in [17] (the torsion is ${}^t b + b$ and it has the same signature as ${}^t b^{-1} + b^{-1} = {}^t b^{-1}({}^t b + b) b^{-1}$).

1.2.2 Tonelli Hamiltonians

A C^2 function $H : \mathbb{T}^* M \to \mathbb{R}$ is a *Tonelli Hamiltonian* if it is:

– superlinear in the fiber, i.e., $\forall A \in \mathbb{R}, \exists B \in \mathbb{R}, \forall (q,p) \in T^*M, \|p\| \geq B \Rightarrow H(q,p) \geq A\|p\|$;

– C^2-convex in the fiber, i.e., for every $(q,p) \in T^*M$, the Hessian $\frac{\partial^2 H}{\partial p^2}$ of H in the fiber direction is positive definite as a quadratic form.

We denote the Hamiltonian flow of H by (φ_t^H) and the Hamiltonian vector-field by X_H. Note that the flow of a Tonelli Hamiltonian defined on $\mathbb{A}_n$ is not necessarily a globally positive diffeomorphism. A geodesic flow is an example of a Tonelli flow. For example, the flat metric on $\mathbb{T}^n$ corresponds to the Tonelli Hamiltonian $H_0(q,p) = \frac{1}{2}\|p\|^2$ and its time-one flow is nothing but the map f_0 that we defined in Example 1.4.

At the end of the 1980s, J. Mather extended Aubry–Mather theory to the Tonelli Hamiltonians, introducing the concept of globally minimizing orbits and minimizing measures (see [23] and [21]).

To explain that, we associate to any Tonelli Hamiltonian $H : T^*M \to \mathbb{R}$ its Lagrangian $L : TM \to \mathbb{R}$ that is dual to H via the formula

$$\forall (q,v) \in TM, L(q,v) = \sup_{p \in T_q^*M} (p.v - H(q,p)).$$

Then L is as regular as H is and is superlinear and C^2-convex in the fiber direction (see for example, [11]). Moreover, we have:

$$L(q,v) + H(q,p) = p.v \Longleftrightarrow v = \frac{\partial H}{\partial p}(q,p) \Longleftrightarrow p = \frac{\partial L}{\partial v}(q,v).$$

If $\gamma : [\alpha, \beta] \to M$ is an absolutely continuous arc, its *Lagrangian action* is then:

$$A_L(\gamma) = \int_\alpha^\beta L(\gamma(t), \dot\gamma(t))dt.$$

1.3 Minimizing measures

1.3.1 Case of the globally positive diffeomorphisms of the two-dimensional annulus

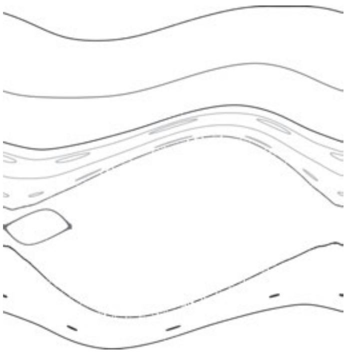
We use the notations that were introduced in subsection 1.2.1. In the two-dimensional case, J. Mather and Aubry & Le Daeron proved in [6] and [22] the existence of orbits $(q_i, p_i)_{i\in\mathbb{Z}}$ for F that are *globally minimizing*. This means that for every $\ell \in \mathbb{Z}$ and every $k \geq 2$, $(q_{\ell+1}, \dots, q_{\ell+k-1})$ is minimizing the function $\mathcal{F}$ defined by:

$$\mathcal{F}(q_{\ell+1}, \dots, q_{\ell+k-1}) = \sum_{i=\ell+1}^{k} S(q_{i-1}, q_i).$$

Then each of these orbits $(q_i, p_i)_{i\in\mathbb{Z}}$ is supported in the graph of a Lipschitz map defined on a closed subset of $\mathbb{T}$, and there exists a bi-Lipschitz orientation preserving homeomorphism $h : \mathbb{T} \to \mathbb{T}$ such that $(q_i)_{i\in\mathbb{Z}} = (h^i(q_0))_{i\in\mathbb{Z}}$. Hence, each of these orbits has a *rotation number*.

Moreover, for each rotation number $\rho \in \mathbb{R}$, there exists a minimizing orbit that has this rotation number and there even exists a *minimizing measure*, i.e., an invariant measure whose support is compact and filled by globally minimizing orbit, such that all the orbits contained in the support have the same rotation number ρ. These supports, which are Lipschitz graphs above a subset of $\mathbb{T}$, are sometimes called *Aubry–Mather set*.

In the following figure that concerns the so-called standard twist map, you can observe some invariant curves, some Cantor subsets, and some periodic islands that must contain one periodic point.



Different kinds of Aubry–Mather sets can occur in this setting:
(1) some of them are invariant loops that are the graphs of some Lipschitz maps $\eta : \mathbb{T} \to \mathbb{R}$;
(2) some other ones are just periodic orbits; and
(3) some of these Aubry–Mather sets are Cantor sets.

In Case 1, it can happen that the dynamics restricted to the curve is bi-Lipschitz conjugate to a rotation; in this case, the Lyapunov exponents of the invariant measure supported in the curve are zero. This is the case for the KAM curves. But P. Le Calvez proved in [18] that in general (i.e., for a dense and G_δ subset of the set of the symplectic twist maps), there exists an open and dense subset U of $\mathbb{R}$ such that any Aubry–Mather set that has its rotation number in U is uniformly hyperbolic.

1.3.2 Case of the globally positive diffeomorphisms in higher dimension

For globally positive diffeomorphism in higher dimension, Garibaldi & Thieullen prove the existence of globally minimizing orbits and measures in [13]. The results that they obtain are very similar to the ones that we recall in Section 1.3.3 for Tonelli Hamiltonians.

Remark 1.5. There also exists an Aubry–Mather theory for time-one maps of time-dependent Tonelli Hamiltonians (see for example [7]). Even when the manifold M is $\mathbb{T}^n$, the time-one map is not necessarily a globally positive diffeomorphism of $\mathbb{A}_n$. Moreover, except for the two-dimensional annulus (see [24]), it is unknown if a globally positive diffeomorphism is always the time-one map of a time-dependent Tonelli Hamiltonian (see Theorem 41.1 in [14] for some partial results). In this survey, we will not speak about these time-one maps.

1.3.3 Case of the Tonelli Hamiltonians

It can be proved that $q_b, q_e \in M$ are two points of M and $\beta > \alpha$ are two real numbers, if $\Gamma(q_b, q_e; \alpha, \beta)$ is the set of the C^2-arcs $\gamma : [\alpha, \beta] \to M$ that join q_b to q_e endowed with the C^2-topology, then γ is a critical point of the restriction of A_L to $\Gamma(q_b, q_e; \alpha, \beta)$ if and only if γ is the projection of an arc of orbit for H. This arc of orbit is then $(\gamma(t), \frac{\partial L}{\partial v}(\gamma(t), \dot\gamma(t)))_{t\in[\alpha,\beta]}$.

In [23], J. Mather proves the existence of complete orbits $(\varphi_t^H(q,p))_{t\in\mathbb{R}} = (q(t), p(t))_{t\in\mathbb{R}}$ that are *globally minimizing*, i.e., such that every arc $(\pi \circ \varphi_t^H(x))_{t\in[\alpha,\beta]} = (q(t))_{t\in[\alpha,\beta]}$ is minimizing for the restriction of A_L to $\Gamma(q(\alpha), q(\beta); \alpha, \beta)$. He also proves the existence of *minimizing measures*, i.e., invariant measures whose support is filled by globally minimizing orbit.

When replacing L by $L+\lambda$, where λ is any closed 1-form on M, we obtain the same critical points for the Lagrangian action $A_{L+\lambda}$ as for the Lagrangian action A_L. But the minima for those two functions are not the same. Hence, adding different closed 1-form λ to L is a way to find other invariant measures supported in graphs, these measures being minimizing for $L + \lambda$. The supports of these measures are the generalization of the Aubry–Mather sets. A rotation number can be associated to any minimizing measure (see [23]) and it can be proved that there exists a minimizing measure for any rotation number. But this does not give the existence of minimizing orbits of any rotation number (indeed the considered measures have not to be ergodic).

2 Lyapunov exponents for the minimizing measures and angle of the Oseledets splitting

Here we are interested in the Lyapunov exponents of the minimizing measures for globally positive diffeomorphisms or Tonelli Hamiltonian flows. In the case of symplectic twist maps, we noticed at the end of Section 1.3.1 that these exponents may be non zero or zero.

Let $(\mathcal{D}_t)$ (t in $\mathbb{Z}$ or $\mathbb{R}$) be either the $\mathbb{Z}$-action generated by a globally positive diffeomorphism or the $\mathbb{R}$-action generated by a Tonelli Hamiltonian. Let μ be an ergodic minimizing measure. A general fact for ergodic measures and C^1-bounded dynamics is that the closer the stable and unstable bundles are (this means that there is an orthogonal basis of the stable bundle that is close to a orthonormal basis of the unstable bundle), the closer to zero the Lyapunov exponents are (see Proposition 2.2 in Section 2.1.1 for a more precise statement and [4] for a proof).

But, in general, the converse assertion is false. We will see that it is true in the case of a twisting dynamics.

- In Section 2.1 , we will prove these two statements for the Dirac measures and even give a precise statement for the first assertion in the general case in Section 2.1.1.
- In Section 2.2 , we will explain that there is link between the number of non zero Lyapunov exponents for a minimizing measure and the dimension of the intersection of the so-called Green bundles.
- In Section 2.3, we will explain the second assertion (and in fact a more precise statement): for a twisting dynamics with an hyperbolic minimizing measure, if the stable and unstable bundles are not close together (this means that all the unit vectors of the stable bundle are far from every unit vector of the unstable bundle), then all the positive Lyapunov exponents are large.

2.1 Some simple remarks for Dirac masses

Before looking at the Lyapunov exponents of any invariant measure, let us have a look into what happens for a Dirac mass in dimension 2.

More precisely, let us assume that x is a fixed point of a two-dimensional diffeomorphism. We assume that $\sup\{\|Df(x)\|, \|(Df(x))^{-1}\|\} \leq C$, where C is some constant. Let λ_1 and λ_2 be the two (complex) eigenvalues for $Df(x)$; then the Lyapunov exponents for the Dirac mass δ_x are $\log(|\lambda_1|)$ and $\log(|\lambda_2|)$. Let us assume that λ_1 and λ_2 are real and let us denote by E_1, E_2 the corresponding eigenspaces.

2.1.1 What happens when the stable and unstable subspaces are close together

We have

Simple principle: If the eigenspaces E_1 and E_2 are close together, then the two eigenvalues λ_1 and λ_2 have to be close together too.

More precisely, if e_i is a unit vector on E_i, we have

$$|\lambda_1 - \lambda_2| \leq 2C \inf\left\{\frac{\|e_2 - e_1\|}{\|e_1 + e_2\|}, \frac{\|e_1 + e_2\|}{\|e_1 - e_2\|}\right\}.$$

Proof of the simple principle. We just compute

$$Df(x).\frac{e_2 - e_1}{\|e_2 - e_1\|} = \frac{\lambda_2 + \lambda_1}{2}.\frac{e_2 - e_1}{\|e_2 - e_1\|} + \frac{\lambda_2 - \lambda_1}{2}\frac{e_2 + e_1}{\|e_2 - e_1\|}.$$

As $e_2 - e_1$ and $e_1 + e_2$ are orthogonal, we deduce

$$\frac{|\lambda_2 - \lambda_1|}{2} \frac{\|e_2 + e_1\|}{\|e_2 - e_1\|} \leq \left\| Df(x). \frac{e_2 - e_1}{\|e_2 - e_1\|} \right\| \leq \|Df(x)\| \leq C.$$

Changing e_1 into $-e_1$, we obtain the second inequality. □

If f is symplectic, we have $\lambda_2 = \frac{1}{\lambda_1}$. In this case, if the eigenspaces are close together, the two eigenvalues have to be close to 1 and then the Lyapunov exponents are close to 0.

This simple remark for fixed point can be generalized to any dimension and any invariant measure in the following way. For a proof, see [4].

Notation 2.1. We endow a compact manifold N with a Riemannian metric. If E, F are two linear subspaces of T_xN that are d-dimensional with $d \geq 1$, the distance between E and F is

$$\mathrm{dist}(E, F) = \inf_{(e_i),(f_i)} \max\{\|e_1 - f_1\|, \dots, \|e_d - f_d\|\},$$

where the infimum is taken over all the orthonormal basis (e_i) of E, (f_i) of F.

Proposition 2.2. *Let K be a compact subset of a manifold N, let $C > 0$ be a real number. Then, for any $f \in \mathrm{Diff}^1(M)$ so that $\max\{\|Df_{|K}\|, \|Df^{-1}_{|K}\|\} \leq C$, if f has an invariant ergodic measure μ with support in K such that the Oseledets stable and unstable bundles E^s and E^u of μ have the same dimension d , if we denote by Λ_u the sum of the positive Lyapunov exponents and by Λ_s the sum of the negative Lyapunov exponents, then:*

$$0 < \Lambda_u - \Lambda_s \leq d \log\left(1 + (C^2 + 1) \int \mathrm{dist}(E^u, E^s) d\mu\right),$$

where dist *is the distance.*

If, for example, f is a symplectic diffeomorphism of T^*M, then E^u and E^s have same dimension (see for example, [9]). We deduce from the above proposition that if the stable and unstable Oseledets bundles are close together, then all the Lyapunov exponents are close to 0. This result is not very surprising and not specific to the twisting dynamics. What is more surprising and specific to the twisting dynamics will come in subsection 2.1.2.

2.1.2 What happens when the stable and unstable subspaces are far from each other

For general symplectic dynamics, we can have simultaneously two eigenvalues that are close together and two eigenspaces that are not close together. See, for example, the linear isomorphism of $\mathbb{R}^2$ with matrix in the usual basis

$$M = \begin{pmatrix} 1+\varepsilon & 0 \\ 0 & \frac{1}{1+\varepsilon} \end{pmatrix},$$

with $\varepsilon > 0$ small enough.

As noticed by J.-C. Yoccoz, this cannot happen for the minimizing Dirac masses of a twist map of the two-dimensional annulus. It can be proved that at the fixed point x corresponding to such a minimizing Dirac mass δ_x, the eigenvalues of $Df(x)$ are real. We denote them by λ_1, λ_2 and by E_1, E_2 the two corresponding eigenspaces and by $M = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ the matrix of $Df(x)$ in coordinates (q, p). Then we have

Simple result If the torsion b is bounded from below by a positive number, if E_1 and E_2 are far from each other, then $|\lambda_2 - \lambda_1|$ cannot be too small.

More precisely, if θ is the angle between E_1 and E_2, then we have

$$|b| \leq \sup\{1, \left(\operatorname{cotan}(\theta)\right)^2\}|\lambda_2 - \lambda_1|.$$

Proof of the simple result. The angle between E_1 and E_2 being θ, there exists a matrix R of rotation such that if $P := R\left(\begin{smallmatrix} 1 & \operatorname{cotan}\theta \\ 0 & 1 \end{smallmatrix}\right)$, then $P\left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right) = e_1$ and $P\left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right) \in \mathbb{R}.e_2$. As R is a matrix of rotation, the modulus of all the coefficients of $P = \left(\begin{smallmatrix} \alpha & \beta \\ \delta & \gamma \end{smallmatrix}\right)$ is less than $\sup\{1, |\operatorname{cotan}\theta|\}$. Moreover, we have

$$M = \begin{pmatrix} \gamma & -\beta \\ -\delta & \alpha \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \cdot \begin{pmatrix} \alpha & \beta \\ \delta & \gamma \end{pmatrix} = \begin{pmatrix} * & \delta.\gamma(\lambda_1 - \lambda_2) \\ * & * \end{pmatrix}.$$

We deduce that $b = \delta\beta(\lambda_1 - \lambda_2)$ and the wanted result. □

2.2 Number of non zero Lyapunov exponents

Before giving an estimation of the non zero Lyapunov exponents, we will try to find how many they are. As the dynamics is symplectic, we know that the number of negative Lyapunov exponents is equal to the number of positive Lyapunov exponents and then the number of zero Lyapunov exponents is even (see [9] for a proof).

2.2.1 The two Green bundles

The Green bundles are two Lagrangian bundles that are defined along the minimizing orbits. In general, they are measurable but not continuous. Let us recall that a subspace H of the symplectic space $T_x(T^*M)$ is Lagrangian if its dimension is n and if the restriction of the symplectic form to H vanishes: $\omega_{|H\times H} = 0$.

The Green bundles were introduced in the 1950s by L. W. Green to give a proof of the two-dimensional version of Hopf conjecture: a Riemannian metric of $\mathbb{T}^n$ with no conjugate points is flat. Then P. Foulon extended the construction to the Finsler metrics in [12] and G. Contreras & R. Iturriaga built them for any Tonelli Hamiltonian in [10]. The construction for the twist maps of the annulus, and more generally for the twist maps of $\mathbb{T}^n \times \mathbb{R}^n$ is due to M. Bialy & R. MacKay (see [8]).

We will recall here their precise definition and we will give their main properties. Before this, let us recall that there exists a way to compare different Lagrangian subspaces of $T_x(T^*M)$ that are transverse to the vertical $V(x)$. We choose some coordinates (q, p) as explained at the beginning of Section 1 and we denote the linearized coordinates by $(\delta q, \delta p)$ of $T_x(T^*M)$. If H_1, H_2 are two Lagrangian subspaces of $T_x(T^*M)$ that are transverse to the vertical $V(x)$, we can write them in coordinates $(\delta q, \delta p)$ as the graph of some symmetric matrices S_1, S_2. We say that L_1 is under L_2 and write $L_1 \le L_2$ when $S_2 - S_1$ is a positive semi-definite matrix. We say that L_1 is strictly under L_2 and write $L_1 < L_2$ if $L_1 \le L_2$ and L_1 and L_2 are transverse. This is equivalent to say that $S_2 - S_1$ is positive definite. It can be proved that this definition does not depend on the chart that we choose. For an equivalent but more intrinsic definition, see [1].

Along every minimizing orbit of a globally positive diffeomorphism $F : \mathbb{T}^n \times \mathbb{R}^n \to \mathbb{T}^n \times \mathbb{R}^n$ or a Tonelli Hamiltonian flow $H : T^*M \to \mathbb{R}$ that we will denote by $(\mathcal{D}_t)$ (with t in $\mathbb{Z}$ or $\mathbb{R}$), we can define two Lagrangian bundles G_- and G_+.

Definition 2.3. If the orbit of x is minimizing, then the family $(D\mathcal{D}_t.V(\mathcal{D}_{-t}x))_{t>0}$ is a decreasing family of Lagrangian subspaces that converges to $G_+(x)$ and the familly $(D\mathcal{D}_{-t}.V(\mathcal{D}_t x))_{t>0}$ is an increasing family of Lagrangian subspaces that converges to $G_-(x)$.

We recall now some properties of the Green bundles:
- they are transverse to the vertical and $G_- \le G_+$;
- G_- and G_+ are invariant by the linearized dynamics, i.e., $D\mathcal{D}_t.G_\pm = G_\pm \circ \mathcal{D}_t$;
- for every compact K such that the orbit of every point of K is minimizing, the two Green bundles restricted to K are uniformly far from the vertical;
- (dynamical criterion) if the orbit of x is minimizing and relatively compact in T^*M, if $\liminf_{t\to+\infty} \|D(\pi \circ \mathcal{D}_t)(x)v\| < +\infty$ then $v \in G_-(x)$.
  If $\liminf_{t\to+\infty} \|D(\pi \circ \mathcal{D}_{-t})(x)v\| < +\infty$ then $v \in G_+(x)$.

The bundles G_- and G_+ are the *Green bundles*. The proof of the results that we mentioned before can be found in [1] for the Tonelli Hamiltonians and in [4] for the globally positive diffeomorphisms.

An easy consequence of the dynamical criterion and the fact that the Green bundles are Lagrangian is that when there is a splitting of $T_x(T^*M)$ into the sum of a stable, a center and a unstable bundles $T_x(T^*M) = E^s(x) \oplus E^c(x) \oplus E^u(x)$, for example an Oseledets splitting or a partially hyperbolic splitting, then we have

$$E^s \subset G_- \subset E^s \oplus E^c \quad \text{and} \quad E^u \subset G_+ \subset E^u \oplus E^c.$$

Let us give the argument of the proof. Because of the dynamical criterion, we have $E^s \subset G_-$. Because the dynamical system is symplectic, the symplectic orthogonal subspace to E^s is $(E^s)^\perp = E^s \oplus E^c$ (see e.g., [9]). Because G_- is Lagrangian, we have $G_-^\perp = G_-$. We obtain then $G_-^\perp = G_- \subset E^{s\perp} = E^s \oplus E^c$.

Let us note the following straightforward consequence: for a minimizing measure, the whole information concerning the positive (resp. negative) Lyapunov exponents is contained in the restricted linearized dynamics $D\mathcal{D}_{t|G_+}$ (resp. $D\mathcal{D}_{t|G_-}$). In particular, when the measure is weakly hyperbolic, we have almost everywhere $G_+ = E^u$ and $G_- = E^s$.

Notation 2.4. Using a Riemannian metric on M, we define the horizontal subspace $\mathcal{H}$ as the kernel of the connection map. Then, for every Lagrangian subspace $\mathcal{G}$ of $T_x(T^*M)$, there exists a linear map $G : \mathcal{H}(x) \to V(x)$ whose graph is $\mathcal{G}$. That is the meaning of *graph* in what follows. When $M = \mathbb{T}^n$, we choose of course $\mathcal{H} = \mathbb{R}^n \times \{0\}$.

We denote by s_+ (resp. s_-) the linear map $\mathcal{H} \to V$ with graph G_+ (resp. G_-). When we use symplectic coordinates, their matrices are symmetric.

Along a minimizing orbit in the case of a globally positive diffeomorphism, $G_k = Df^k(V \circ f^{-k})$ (resp. $G_{-k} = Df^{-k}(V \circ f^k)$) is the graph of s_k (resp. s_{-k}).

2.2.2 Link between the central dimension and the dimension of $G_- \cap G_+$

From $E^s \subset G_- \subset E^s \oplus E^c$ and $E^u \subset G_+ \subset E^u \oplus E^c$, we deduce that $G_- \cap G_+ \subset E^c$. Hence, $G_- \cap G_+$ is an isotropic subspace (for ω) of the symplectic space E^c. We deduce that $\dim(E^c) \geq 2\dim(G_- \cap G_+)$. When $E^s \oplus E^c \oplus E^u$ designates the Oseledets splitting of some minimizing measure μ , what is proved in [3] is that this inequality is an equality μ almost everywhere for the Tonelli Hamiltonian flows and the same result is proved for the globally positive diffeomorphisms in [5].

Theorem 2.5. *Let $(\mathcal{D}_t)$ (t in $\mathbb{Z}$ or $\mathbb{R}$) be either the $\mathbb{Z}$-action generated by a globally positive diffeomorphism or the $\mathbb{R}$-action generated by a Tonelli Hamiltonian. Let μ be a minimizing measure and let us denote by p the μ-almost everywhere dimension of $G_- \cap G_+$. Then μ has exactly $2p$ zero Lyapunov exponents, $n - p$ positive Lyapunov exponents, and $n - p$ negative Lyapunov exponents.*

The idea is the following one. Firstly, let us notice that we have nothing to prove when $\dim(G_- \cap G_+) = n$ because we know that $\dim(E^c) \geq 2\dim(G_- \cap G_+) = 2n$; in this case, $\dim(E^c) = 2n$ and all the Lyapunov exponents are zero.

In the other case, we consider the following restricted-reduced linearized dynamics. Let μ be an ergodic minimizing measure. Then the quantity $\dim(G_- \cap G_+)$ is μ-almost everywhere constant. We denote this dimension by p and we assume that $p < n$.

Notation 2.6. We introduce the following linear spaces (see [3]): $E = G_- + G_+$, $R = G_- \cap G_+$, and F is the reduced space $F = E/R$. As E is coisotropic for ω with $E^{\perp\omega} = R$, then F is the symplectic reduction of E. As E and R are invariant by the linearized dynamics, then we can define a cocycle M_t on F as the reduced linearized dynamics. This cocycle is then symplectic for the reduced symplectic form Ω.

In [3] and [5], we define for the cocycle (M_t) a vertical subspace, some reduced Green bundles g_- and g_+ that have properties similar to the ones of $G_\pm$, and we prove that g_- and g_+ are transverse μ-almost everywhere. As we will explain in subsection 2.2.3, the transversality of the Green bundles implies the (weak) hyperbolicity of the measure. Here we have only the transversality of the reduced Green bundles, but this imply that the cocycle (M_t) is (weakly) hyperbolic and then that the linearized dynamics has at least $2(n-p)$ non zero Lyapunov exponents. This gives the conclusion.

2.2.3 The transversality of the two Green bundles implies some hyperbolicity

We will explain here why a minimizing measure μ is weakly hyperbolic when the Green bundles are transverse almost everywhere. We will deal with the discrete case (i.e., globally positive diffeomorphisms of $\mathbb{A}_n$). The diffeomorphism is denoted by f and we assume that μ-almost everywhere we have $T_x\mathbb{A}_n = G_-(x) \oplus G_+(x)$. We want to prove that f has at least n positive Lyapunov exponents; in this case, because f is symplectic, μ has also n negative Lyapunov exponents (see [9]).

The idea is to use a bounded (but noncontinuous) symplectic change of linearized coordinates along the minimizing orbits where $T_x\mathbb{A}_n = G_-(x) \oplus G_+(x)$ such that G_+ becomes the horizontal and that preserves the vertical space. Because G_+ is invariant by Df, the symplectic matrix of Df^k is $M^k(x) = \begin{pmatrix} a_k(x) & b_k(x) \\ 0 & d_k(x) \end{pmatrix}$.

Because G_k is transverse to the vertical, we have $\det b_k \neq 0$. Because of the definition of G_k, we have then $d_k(x) = s_k(f^k x)b_k(x)$. As $(s_k(x))_{k\geq 1}$ is decreasing and tends to $\mathbf{0}$ (because the horizontal is G_+), the symmetric matrix $s_k(f^k x)$ is positive definite. Moreover, because the matrix $M^k(x)$ is symplectic, we have

$$\left(M^k(x)\right)^{-1} = \begin{pmatrix} {}^t d_k(x) & -{}^t b_k(x) \\ 0 & {}^t a_k(x) \end{pmatrix},$$

and by definition of $G_{-k}(x)$: ${}^t a_k(x) = -s_{-k}(x){}^t b_k(x)$, and finally we have

$$M^k(x) = \begin{pmatrix} -b_k(x)s_{-k}(x) & b_k(x) \\ 0 & s_k(f^k x)b_k(x) \end{pmatrix}.$$

The proof is then made of several lemmata. The first one is a consequence of Egorov theorem and of the fact that μ-almost everywhere on $\mathrm{supp}\mu$, G_+ and G_- are transverse and then $-s_-$ is positive definite.

Lemma 2.7. *For every $\varepsilon > 0$, there exists a measurable subset J_ε of* $\mathrm{supp}\mu$ *such that*
- $\mu(J_\varepsilon) \geq 1 - \varepsilon$;
- *on J_ε, $(s_k)_{k\geq 1}$ and $(s_{-k})_{k\geq 1}$ converge uniformly;*
- *there exists a constant $\alpha = \alpha(\varepsilon) > 0$ such that $\forall x \in J_\varepsilon$, $-s_-(x) \geq \alpha \mathbf{1}$.*

We deduce:

Lemma 2.8. *Let J_ε be as in the previous lemma. On the set $\{(k, x) \in \mathbb{N} \times J_\varepsilon, f^k(x) \in J_\varepsilon\}$, the sequence of conorms $(m(b_k(x))$ converges uniformly to $+\infty$, where $m(b_k) = \|b_k^{-1}\|^{-1}$.*

Proof. Let k, x be as in the lemma.

The matrix $M_k(x) = \begin{pmatrix} -b_k(x)s_{-k}(x) & b_k(x) \\ 0 & s_k(f^k x)b_k(x) \end{pmatrix}$ being symplectic, we have $-s_{-k}(x)^t b_k(x)s_k(f^k x)b_k(x) = \mathbf{1}$ and thus $-b_k(x)s_{-k}(x)^t b_k(x)s_k(f^k x) = \mathbf{1}$ and $b_k(x)s_{-k}(x)^t b_k(x) = -\left(s_k(f^k x)\right)^{-1}$.

We know that on J_ε, (s_k) converges uniformly to zero. Hence, for every $\delta > 0$, there exists $N = N(\delta)$ such that $k \geq N \Rightarrow \|s_k(f^k x)\| \leq \delta$.

Moreover, as $G_{-1} \leq G_{-k} \leq G_1$ and G_{-1} and G_1 continuously depend on x in the compact subset $\mathrm{supp}\mu$ and because the linear change of coordinates that we use is bounded, there exists $\beta > 0$ so that $\|s_{\pm k}\| \leq \beta$ uniformly in k on $\mathrm{supp}\mu$. Hence, if we choose $\delta' = \frac{\delta^2}{\beta}$, for every $k \geq N = N(\delta')$ and $x \in J_\varepsilon$ such that $f^k x \in J_\varepsilon$, we obtain

$$\begin{aligned} \forall v \in \mathbb{R}^p, \beta \|{}^t b_k(x)v\|^2 &= {}^t v b_k(x)(\beta \mathbf{1})^t b_k(x)v \\ &\geq -{}^t v b_k(x)s_{-k}(x)^t b_k(x)v = {}^t v \left(s_k(f^k x)\right)^{-1} v, \end{aligned}$$

and we have: ${}^t v(s_k(f^k x))^{-1} v \geq \frac{\beta}{\delta^2}\|v\|^2$ because $s_k(f^k x)$ is a positive definite matrix that is less than $\frac{\delta^2}{\beta}\mathbf{1}$. We finally obtain $\|{}^t b_k(x)v\| \geq \frac{1}{\delta}\|v\|$ and then the result that we wanted. □

Notation 2.9. We choose $\beta > 0$ as in the previous proof, i.e., such that $\forall k \in \mathbb{Z}\backslash\{0\}, \forall x \in \mathrm{supp}\mu, \|s_k(x)\| \leq \beta$.

From now we fix a small constant $\varepsilon > 0$, associate a set J_ε with ε via Lemma 2.7 and a constant $0 < \alpha < \beta$; then there exists $N \geq 0$ such that

$$\forall x \in J_\varepsilon, \forall k \geq N, f^k(x) \in J_\varepsilon \Rightarrow m(b_k(x)) \geq \frac{2}{\alpha}.$$

Lemma 2.10. *Let J_ε be as in Lemma 2.7. For μ-almost point x in J_ε, there exists a sequence of integers $(j_k) = (j_k(x))$ tending to $+\infty$ such that*

$$\forall k \in \mathbb{N}, m(b_{j_k}(x)s_{-j_k}(x)) \geq \left(2^{\frac{1-\varepsilon}{2N}}\right)^{j_k}.$$

Proof. As μ is ergodic for f, we deduce from Birkhoff ergodic theorem that for almost every point $x \in J_\varepsilon$, we have

$$\lim_{\ell\to+\infty} \frac{1}{\ell}\sharp\{0 \leq k \leq \ell - 1; f^k(x) \in J_\varepsilon\} = \mu(J_\varepsilon) \geq 1 - \varepsilon.$$

We introduce the notation $N(\ell) = \sharp\{0 \leq k \leq \ell - 1; f^k(x) \in J_\varepsilon\}$.

For such an x and every $\ell \in \mathbb{N}$, we find a number $n(\ell)$ of integers

$$0 = k_1 \leq k_1 + N \leq k_2 \leq k_2 + N \leq k_3 \leq k_3 + N \leq \cdots \leq k_{n(\ell)} \leq \ell,$$

such that $f^{k_i}(x) \in J_\varepsilon$ and $n(\ell) \geq \left[\frac{N(\ell)}{N}\right] \geq \frac{N(\ell)}{N} - 1$. In particular, we have $\frac{n(\ell)}{\ell} \geq \frac{1}{N}\left(\frac{N(\ell)}{\ell} - \frac{N}{\ell}\right)$, the right term converging to $\frac{\mu(J_\varepsilon)}{N} \geq \frac{1-\varepsilon}{N}$ when ℓ tends to $+\infty$. Hence, for ℓ large enough, we find $n(\ell) \geq 1 + \ell\frac{1-\varepsilon}{2N}$.

As $f^{k_i}(x) \in J_\varepsilon$ and $k_{i+1} - k_i \geq N$, we have $m(b_{k_{i+1}-k_i}(f^{k_i}(x))) \geq \frac{2}{\alpha}$. Moreover, we have $s_{-(k_{i+1}-k_i)}(f^{k_i}x) \leq s_-(f^{k_i}x) \leq -\alpha\mathbf{1}$ then $m(s_{-(k_{i+1}-k_i)}(f^{k_i}x)) \geq \alpha$; hence,

$$m(b_{k_{i+1}-k_i}(f^{k_i}x)s_{-(k_{i+1}-k_i)}(f^{k_i}x)) \geq 2.$$

But the matrix $-b_{k_{n(\ell)}}(x)s_{-k_{n(\ell)}}(x)$ is the product of $n(\ell) - 1$ such matrices. Hence,

$$m(b_{k_{n(\ell)}}(x)s_{-k_{n(\ell)}}(x)) \geq 2^{n(\ell)-1} \geq 2^{\ell\frac{1-\varepsilon}{2N}} \geq \left(2^{\frac{1-\varepsilon}{2N}}\right)^{k_{n(\ell)}}.$$

□

This implies that all the Lyapunov exponents of the restriction of Df to G_+ are greater than $\log\left(2^{\frac{1-\varepsilon}{2N}}\right) > 0$.

2.3 Lower bounds for the positive Lyapunov exponents

Notation 2.11. For a positive semi-definite symmetric matrix S that is not the zero matrix, we denote by $q_+(S)$ its smallest positive eigenvalue.

Theorem 2.12. *Let μ be an ergodic minimizing measure of a globally positive diffeomorphism of $\mathbb{A}_n$ that has at least one non zero Lyapunov exponent.We denote the smallest positive Lyapunov exponent of μ by $\lambda(\mu)$ and an upper bound for $\|s_1 - s_{-1}\|$ above* suppμ *by C. Then we have*

$$\lambda(\mu) \geq \frac{1}{2}\int \log\left(1 + \frac{1}{C}q_+((s_+ - \mathbb{S})(x))\right) d\mu(x).$$

The proof of this result is given in [5]. There is a similar result for Tonelli Hamiltonians.

Theorem 2.13. *Let μ be an ergodic minimizing measure for a Tonelli Hamiltonian $H : T^*M \to \mathbb{R}$ and with at least one non zero Lyapunov exponent; then its smallest positive Lyapunov exponent $\lambda(\mu)$ satisfies $\lambda(\mu) \geq \frac{1}{2}\int m(\frac{\partial^2 H}{\partial p^2}) \cdot q_+(s_+ - \mathbb{S})d\mu$.*

The proof of Theorem 2.12 is a little long and involves some technical changes of bases. We prefer to give the proof of Theorem 2.13, that is simpler and shorter. The first point is the following lemma.

Lemma 2.14. *Let $H : T^*M \to \mathbb{R}$ be a Tonelli Hamiltonian. Let (x_t) be a minimizing orbit and let U and S be two Lagrangian bundles along this orbit that are invariant by the linearized Hamilton flow and transverse to the vertical. Let $\delta x_U \in U$ be an infinitesimal orbit contained in the bundle U and let us denote by δx_S the unique vector of S such that $\delta x_U - \delta x_S \in V$ (hence δx_S is* not *an infinitesimal orbit). Then*

$$\frac{d}{dt}(\omega(x_t)(\delta x_S(t), \delta x_U(t))) = {}^t(\delta x_U(t) - \delta x_S(t))\frac{\partial^2 H}{\partial p^2}(x_t)(\delta x_U(t) - \delta x_S(t)) \geq 0.$$

Proof. As the result that we want to prove is local, we can assume that we are in the domain of a dual chart and express all the things in the corresponding dual linearized coordinates.

We consider an invariant Lagrangian linear bundle G that is transverse to the vertical along the orbit of $x = (q, p)$. We denote the symmetric matrix whose graph is G by G again. An infinitesimal orbit contained in this bundle satisfies $\delta p = G\delta q$. We deduce from the linearized Hamilton equations (if we are along the orbit $(q(t), p(t)) = x(t)$, $\dot{G}$ designates $\frac{d}{dt}(G(x(t)))$) that

$$\delta\dot{q} = \left(\frac{\partial^2 H}{\partial q \partial p} + \frac{\partial^2 H}{\partial p^2}G\right)\delta q;$$
$$\delta\dot{p} = \left(\dot{G} + G\frac{\partial^2 H}{\partial q \partial p} + G\frac{\partial^2 H}{\partial p^2}G\right)\delta q = -\left(\frac{\partial^2 H}{\partial q^2} + \frac{\partial^2 H}{\partial p \partial q}G\right)\delta q.$$

We deduce from these equations the classical Ricatti equation (it is given for example in [10] for Tonelli Hamiltonians, but the reader can find the initial and simpler Ricatti equation given by Green in the case of geodesic flows in [15]):

$$\dot{G} + G\frac{\partial^2 H}{\partial p^2}G + G\frac{\partial^2 H}{\partial q \partial p} + \frac{\partial^2 H}{\partial p \partial q}G + \frac{\partial^2 H}{\partial p^2} = 0.$$

Let us assume now that the graphs of the symmetric matrices $\mathbb{U}$ and $\mathbb{S}$ are invariant by the linearized flow along the same orbit. We denote by $(\delta q_U, \mathbb{U}\delta q_U)$ an infinitesimal orbit that is contained in the graph of $\mathbb{U}$. Then we have

$$\begin{aligned}
\frac{d}{dt}({}^t\delta q_U(\mathbb{U}-\mathbb{S})\delta q_U) &= 2{}^t\delta q_U(\mathbb{U}-\mathbb{S})\delta\dot{q}_U + {}^t\delta q_U(\dot{\mathbb{U}}-\dot{\mathbb{S}})\delta q_U \\
&= 2{}^t\delta q_U(\mathbb{U}-\mathbb{S})\left(\frac{\partial^2 H}{\partial q\partial p} + \frac{\partial^2 H}{\partial p^2}\mathbb{U}\right)\delta q_U \\
&\quad + {}^t\delta q_U\left(\mathbb{S}\frac{\partial^2 H}{\partial p^2}\mathbb{S} - \mathbb{U}\frac{\partial^2 H}{\partial p^2}\mathbb{U} + \mathbb{S}\frac{\partial^2 H}{\partial q\partial p}\right. \\
&\quad \left. + \frac{\partial^2 H}{\partial p\partial q}\mathbb{S} - \mathbb{U}\frac{\partial^2 H}{\partial q\partial p} - \frac{\partial^2 H}{\partial p\partial q}\mathbb{U}\right)\delta q_U \\
&= {}^t\delta q_U\left(\mathbb{U}\frac{\partial^2 H}{\partial q\partial p} - \mathbb{S}\frac{\partial^2 H}{\partial q\partial p} + \mathbb{U}\frac{\partial^2 H}{\partial p^2}s_+ - 2\mathbb{S}\frac{\partial^2 H}{\partial p^2}\mathbb{U}\right. \\
&\quad \left. + \mathbb{S}\frac{\partial^2 H}{\partial p^2}\mathbb{S} + \frac{\partial^2 H}{\partial p\partial q}\mathbb{S} - \frac{\partial^2 H}{\partial p\partial q}\mathbb{U}\right)\delta q_U \\
&= {}^t\delta q_U(\mathbb{U}-\mathbb{S})\frac{\partial^2 H}{\partial p^2}(\mathbb{U}-\mathbb{S})\delta q_U \geq 0.
\end{aligned}$$

To finish the proof, we just need to notice that in coordinates $\omega(\delta x_S, \delta x_U) =$

$$\begin{aligned}
\omega(\delta x_U, \delta x_U - \delta x_S) &= {}^t(\delta q_U, \mathbb{U}\delta q_U)\begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix}\begin{pmatrix} 0 \\ (\mathbb{U}-\mathbb{S})\delta q_U \end{pmatrix} \\
&= {}^t\delta q_U(\mathbb{U}-\mathbb{S})\delta q_U.
\end{aligned}$$ □

Let μ be an ergodic minimizing Borel probability measure for a Tonelli Hamiltonian $H : T^*M \to \mathbb{R}$ and with at least one non zero Lyapunov exponent; its support K is compact and then, there exists a constant $C > 0$ such that s_+ and s_- are bounded by C above K. We choose a point (q, p) that is generic for μ and $\delta x_+ = (\delta q, s_+\delta q)$ in the Oseledets bundle corresponding to the smallest positive Lyapunov exponent $\lambda(\mu)$ of μ and we introduce $\delta x_- = (\delta q, s_-\delta q)$. Using the linearized Hamilton equations (see Lemma 2.14), because $\omega(x_t)(\delta x_-, \delta x_+) = {}^t\delta q(s_+ - s_-)\delta q$, we obtain

$$\frac{d}{dt}({}^t\delta q(s_+ - s_-)\delta q) = {}^t\delta q(s_+ - s_-)\frac{\partial^2 H}{\partial p^2}(q_t, p_t)(s_+ - s_-)\delta q.$$

Let us notice that $(s_+ - s_-)^{\frac{1}{2}}\delta q$ is contained in the orthogonal space to the kernel of $s_+ - s_-$. Hence

$$\frac{d}{dt}({}^t\delta q(s_+ - s_-)\delta q) \geq m\left(\frac{\partial^2 H}{\partial p^2}\right)q_+(s_+ - s_-){}^t\delta q(s_+ - s_-)\delta q.$$

Moreover, $\delta q \notin \ker(s_+ - s_-)$ because $(\delta q, s_+\delta q)$ corresponds to a positive Lyapunov exponent and then $(\delta q, s_+\delta q) \notin G_- \cap G_+$. Then

$$\begin{aligned}\frac{2}{T}\log(\|\delta q(T)\|) + \frac{\log 2C}{T} &\geq \frac{1}{T}\log({}^t\delta q(T)(s_+ - s_-)(q_T, p_T)\delta q(T))\\ &\geq \frac{1}{T}\log({}^t\delta q(0)(s_+ - s_-)(q, p)\delta q(0))\\ &\quad + \frac{1}{T}\int_0^T m\left(\frac{\partial^2 H}{\partial p^2}(q_t, p_t)\right) q_+((s_+ - s_-)(q_t, p_t))dt.\end{aligned}$$

Using Birkhoff's ergodic theorem, we obtain

$$\lim_{T\to+\infty}\frac{1}{T}\log(\|\delta q(T)\|) = \lambda(\mu) \geq \frac{1}{2}\int m\left(\frac{\partial^2 H}{\partial p^2}\right) q_+(s_+ - s_-)d\mu.$$

3 Shape of the support of the minimizing measures and Lyapunov exponents

3.1 Some notations and definitions

For any subset $A \neq \emptyset$ of a manifold M and any point $a \in A$, different kinds of subsets of T_aM can be defined that are cones and also a generalizations of the notion of tangent space to a submanifold. We introduce them here when $M = \mathbb{R}^n$, but by using some charts, the definition can be extended to any manifold.

Definition 3.1. Let $A \subset \mathbb{R}^n$ a nonempty subset of $\mathbb{R}^n$ and let $a \in A$ be a point of A. Then

- the *contingent cone* to A at a is defined as being the set of all the limit points of the sequences $t_k(a_k - a)$, where (t_k) is a sequence of real numbers and (a_k) is a sequence of elements of A that converges to a. This cone is denoted by C_aA and it is a subset of $T_a\mathbb{R}^n$.
- the *limit contingent cone* to A at a is the set of the limit points of sequences $v_k \in C_{a_k}A$, where (a_k) is any sequence of points of A that converges to a. It is denoted by $\widetilde{C}_aA$ and it is a subset of $T_a\mathbb{R}^n$;
- the *paratangent cone* to A at a is the set of the limit points of the sequences

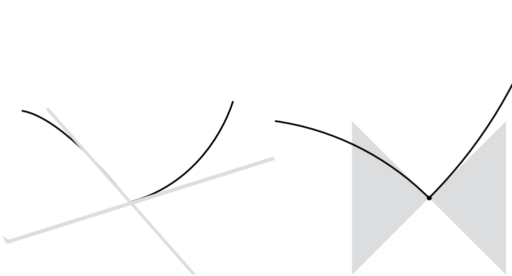
  $$\lim_{k\to\infty} t_k(x_k - y_k),$$

  where (x_k) and (y_y) are sequences of elements of A converging to a and (t_k) is a sequence of elements of $\mathbb{R}$. It is denoted by P_aA and it is a subset of $T_a\mathbb{R}^n$.

The following inclusions are always satisfied,

$$C_aA \subset \widetilde{C}_aA \subset P_aA.$$

Let us give an example of a contingent and paratangent cone at a point where A has an angle.



In sections 3.2, 3.3, and 3.4, we will try to explain some relations between the Green bundles and these tangent cones. Unfortunately, in some cases, we need to use some modified Green bundles (see Section 3.4). In general, the tangent cones are not Lagrangian subspaces (they are neither subspaces nor isotropic). Because we need to compare them to Lagrangian subspaces, we give a definition.

Definition 3.2. Let $\mathcal{L}_- \le \mathcal{L}_+$ be two Lagrangian subspaces of $T_x(T^*M)$ that are transverse to the vertical. If $v \in T_x(T^*M)$ is a vector, we say that v is between $\mathcal{L}_-$ and $\mathcal{L}_+$ and write $\mathcal{L}_- \le v \le \mathcal{L}_+$ if there exists a third Lagrangian subspace in $T_x(T^*M)$ such that
- $v \in \mathcal{L}$;
- $\mathcal{L}_- \le \mathcal{L} \le \mathcal{L}_+$.

A subset B of $T_x(T^*M)$ is between $\mathcal{L}_-$ and $\mathcal{L}_+$ if $\forall v \in B, \mathcal{L}_- \le v \le \mathcal{L}_+$. Then we write $\mathcal{L}_- \le B \le \mathcal{L}_+$.

Remark 3.3. In the two-dimensional case, v is between $\mathcal{L}_-$ and $\mathcal{L}_+$ if and only if the slope of the line generated by v is between the slopes of $\mathcal{L}_-$ and $\mathcal{L}_+$. In higher dimension, it is more complicated.

Definition 3.4.
- A subset A of $\mathbb{R}^n \times \mathbb{R}^n$ is C^1*-isotropic* at some point $a \in A$ if $\widetilde{C}_a A$ is contained in some Lagrangian subspace;
- a subset A of $\mathbb{R}^n \times \mathbb{R}^n$ is C^1*-regular* at some point $a \in A$ if $P_a A$ is contained in some Lagrangian subspace.

Of course, the C^1-regularity of A at a point a implies the C^1-isotropy at the same point. But the converse implication is not true.

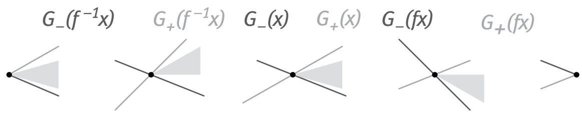
Observe that a C^1 Lagrangian submanifold is always C^1-regular.

3.2 Case I: two-dimensional symplectic twist maps

The results that we explain now for the symplectic twist maps of the two-dimensional annulus are proved in [2].

Theorem 3.5. *Let A be an Aubry–Mather set of a symplectic twist map of the two-dimensional annulus $\mathbb{A}_1$. Then we have*

$$\forall a \in A, G_-(a) \le P_a A \le G_+(a).$$



Corollary 3.6. *Let μ be a minimizing ergodic measure of a symplectic twist map of the two-dimensional annulus. If the Lyapunov exponents of μ are zero, then the support* supp (μ) *of μ is C^1-regular μ-almost everywhere.*

Question 3.7. Is there an example of such an invariant measure with zero Lyapunov exponents such that suppμ is not C^1 at every point of suppμ?

Question 3.8. Is there an example of such an invariant measure with non zero Lyapunov exponents such that suppμ is not uniformly hyperbolic?

Moreover, the following result is also true.

Proposition 3.9. *Let μ be a minimizing ergodic measure of a symplectic twist map of the two-dimensional annulus that has an irrational rotation number. If the Lyapunov exponents of μ are non zero, then the support* supp (μ) *of μ is C^1-irregular μ-almost everywhere.*

We have even the following proposition

Proposition 3.10. *Let μ be a minimizing ergodic measure of a symplectic twist map of the two-dimensional annulus that has an irrational rotation number. If the support* supp(μ) *of μ is C^1-irregular everywhere, then* suppμ *is uniformly hyperbolic.*

Hence, the size of the Lyapunov exponents can be read on the shape of supp(μ). But how can we see in practice this irregularity? For example, if we want to "draw" (with a computer) our irregular (and hyperbolic) Aubry–Mather sets, we can use some sequences of minimizing periodic orbits. But if we look at the pictures of Aubry–Mather sets that exist, we see Cantor sets or curves, but we never see angles of the tangent spaces. That is why the following question was raised by X. Buff.

Question 3.11. (X. Buff) Is it possible (for example by using minimizing periodic orbits) to draw some Aubry–Mather sets with "corners"?

3.3 Case II: invariant Lagrangian graphs of Tonelli Hamiltonians

The proofs of the results we present in this section are given in [1]. We obtain a statement similar to Theorem 3.5 and Corollary 3.6 but no analogue to Proposition 3.9. Indeed, let us consider the following example: (ψ_t) is a geodesic Anosov flow defined on the cotangent bundle $T^*\mathcal{S}$ of a closed surface $\mathcal{S}$. Let $\mathcal{N} = T_1^*\mathcal{S}$ be its unit cotangent bundle, which is a three-manifold invariant by (ψ_t). Then a method due to Mañé (see [20]) allows us to define a Tonelli Hamiltonian H on $T^*\mathcal{N}$ such that the restriction of its flow (φ_t) to the zero section $\mathcal{N}$ is (ψ_t): the Lagrangian L associated with H is defined by $L(q,v) = \frac{1}{2}\|\dot{\psi}(q) - v\|^2$, where $\|.\|$ is any Riemannian metric on $\mathcal{N}$. In this case, the zero section is very regular (even C^∞), but the Lyapunov exponents of every invariant measure with support in $\mathcal{N}$ are non zero (except two, the one corresponding to the flow direction and the one corresponding to the energy direction). Hence, it may happen that some exponents are non zero and the support of the measure is very regular.

Theorem 3.12. *Let $\mathcal{G}$ be a Lipschitz Lagrangian graph that is invariant by the flow of a Tonelli Hamiltonian $H : T^*M \to \mathbb{R}$. Then we have*

$$\forall x \in \mathcal{G}, G_-(x) \le P_x\mathcal{G} \le G_+(x).$$

The following corollary is not proved in [1] but is an easy consequence of Theorems 3.12 and 2.5.

Corollary 3.13. *Let μ be a minimizing ergodic measure for a Tonelli Hamiltonian of T^*M. If the Lyapunov exponents of μ are zero and if the support of μ is a graph above the whole manifold M, then the support* $\mathrm{supp}(\mu)$ *of μ is C^1-regular μ-almost everywhere.*

Question 3.14. Is there an example where such a μ has zero Lyapunov and its support is not C^1 at least one point?

With the hypotheses of Corollary 3.13, if we have further information about the restricted dynamics to $\mathrm{supp}(\mu)$, we can improve the result in the following way.

Proposition 3.15. *Let $\mathcal{G}$ be a Lipschitz Lagrangian graph that is invariant by the flow of a Tonelli Hamiltonian $H : T^*\mathbb{T}^n \to \mathbb{R}$. We assume that for some $T > 0$, the restricted time-T map $\varphi_{T|\mathcal{G}}$ is Lipschitz conjugated to some rotation of $\mathbb{T}^n$. Then $\mathcal{G}$ is the graph of a C^1 function.*

When μ is a minimizing measure with a support smaller than a Lagrangian graph, we do not obtain such a result (even if we have the feeling that it could be true). A fundamental tool to prove the previous results is the following proposition (that is proved in [1]).

Proposition 3.16. *Assume that the orbit of $x \in T^*M$ is globally minimizing for the Tonelli Hamiltonian $H : T^*M \to \mathbb{R}$ and that $\mathcal{L}$ defined on $\mathbb{R}$ is such that*

– *every $\mathcal{L}(t)$ is a Lagrangian subspace of $T_{\varphi_t^H(x)}(T^*M)$ that is transverse to the vertical subbundle;*
– $\forall s, t \in \mathbb{R}, D\varphi_{t-s}^H \mathcal{L}(s) = \mathcal{L}(t)$.

Then we have $\forall t \in \mathbb{R}, G_-(\varphi_t^H(x)) \leq \mathcal{L}(t) \leq G_+(\varphi_t^H(x))$.

Using Proposition 3.16 at any point where the invariant Lagrangian graph $\mathcal{G}$ is differentiable, we deduce a similar inequality for $\mathcal{L}$ being the tangent subspace at such a point. Then using a limit (and the notion of Clarke subdifferential), we deduce Theorem 3.12.

If we could obtain a result similar to Proposition 3.16 for vectors (instead of Lagrangian subspaces), we could deduce a similar statement for all minimizing measures. Hence we raise the following question.

Question 3.17. Let $(\mathcal{D}_t)$ (t in $\mathbb{Z}$ or $\mathbb{R}$) be either the $\mathbb{Z}$-action generated by a globally positive diffeomorphism or the $\mathbb{R}$-action generated by a Tonelli Hamiltonian. Assume that the orbit of $x \in T^*M$ is globally minimizing and that the vector $v \in T_x(T^*M)$ is such that $\forall t, D\varphi_t(v) \notin V(\mathcal{D}_t x)$. Is it true that

$$G_-(x) \leq v \leq G_+(x)?$$

Remark 3.18. Without a lot of change, all the results of subsection 3.3 could be proved for any Lipschitz Lagrangian graph that is invariant by a globally positive diffeomorphism of $\mathbb{T}^n \times \mathbb{R}^n$.

3.4 Case III: Tonelli Hamiltonians and globally positive diffeomorphisms

The results contained in subsection 3.4 come from [3] and [5]. They use in a fundamental way a recent theory called the weak KAM theory that was developed by A. Fathi in [11] in the case of the Tonelli Hamiltonians and by E. Garibaldi & P. Thieullen in [13] in the case of the globally positive diffeomorphisms.

Let us now introduce the modified Green bundles that we will use in this section. We use the constant $c_0 = \frac{\sqrt{13}}{3} - \frac{5}{6}$. We identify $T_x(T^*M)$ to $\mathbb{R}^n \times \mathbb{R}^n$ in such a way that $\{0\} \times \mathbb{R}^n = V(x)$ is the vertical subspace and $\mathbb{R}^n \times \{0\}$ is the horizontal subspace $\mathcal{H}$.

Definition 3.19. We denote by $S_\pm(x) : \mathbb{R}^n \to \mathbb{R}^n$ the linear operator such that $G_\pm(x)$ is the graph of $S_\pm(x)$: $G_\pm(x) = \{(v, S_\pm(x)v); v \in \mathbb{R}^n\}$. Then the *modified Green bundles* $G_\pm$ are defined by

$$\widetilde{G}_-(x) = \{(v, (S_-(x) - c_0(S_+(x) - S_-(x)))v); v \in \mathbb{R}^n\}$$

and

$$\widetilde{G}_+(x) = \{(v, (S_+(x) + c_0(S_+(x) - S_-(x)))v); v \in \mathbb{R}^n\}.$$

Remark 3.20. We have

$$\widetilde{G}_- \le G_- \le G_+ \le \widetilde{G}_+.$$

Moreover, only the two following cases are possible:
- either $\widetilde{G}_-(x)$, $G_-(x)$, $G_+(x)$, $\widetilde{G}_+(x)$ are all distinct;
- or $\widetilde{G}_- = G_- = G_+ = \widetilde{G}_+$.

Theorem 3.21. *Let μ be a minimizing measure for a Tonelli Hamiltonian of T^*M. Then*

$$\forall x \in \operatorname{supp}\mu, \widetilde{G}_-(x) \le \widetilde{C}_x(\operatorname{supp}\mu) \le \widetilde{G}_+(x).$$

Hence, the more irregular $\operatorname{supp}\mu$ is, i.e., the bigger the limit contingent cone is, the more distant $\widetilde{G}_-$ and $\widetilde{G}_+$ (and thus G_- and G_+ too) are from each other and the larger the positive Lyapunov exponents are.

Corollary 3.22. *Let $H : T^*M \to \mathbb{R}$ be a Tonelli Hamiltonian and let μ be an ergodic minimizing probability all of whose Lyapunov exponents are zero. Then, at μ-almost every point of the support* $\operatorname{supp}(\mu)$ *of μ, the set* $\operatorname{supp}(\mu)$ *is C^1-isotropic.*

There are two natural questions that are related to Question 3.17 and that concern also the globally positive diffeomorphims.

Question 3.23. Can we replace $\widetilde{C}_x(\operatorname{supp}\mu)$ by $P_x(\operatorname{supp}\mu)$ in Theorem 3.21 and Theorem 3.26?

Question 3.24. Can we replace $\widetilde{G}_\pm(x)$ by $G_\pm(x)$ in Theorem 3.21 and Theorem 3.26?

If the answer to Question 3.23 is positive, we can replace "C^1-isotropic" by "C^1-regular" in Corollary 3.22 and Corollary 3.27.

For the globally positive diffeomorphisms, we obtain a result only for the so-called strongly minimizing measures (the point is that for Tonelli Hamiltonians, minimizing measures are also strongly minimizing).

Definition 3.25. Let F be a lift of a *globally positive* diffeomorphism f with generating function $S:\mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$. A invariant Borel probability ν on $\mathbb{T}^n\times\mathbb{T}^n$ is *strongly minimizing* if ν is a minimizer in the following formula:

$$\inf_{\mu}\int\limits_{\mathbb{R}^n\times\mathbb{R}^n} S(x,y)d\tilde{\mu}(x,y),$$

where the infimum is taken on the set of the Borel probability measures that are invariant by f and $\tilde{\mu}$ is any lift of μ to a fundamental domain of $\mathbb{R}^n\times\mathbb{R}^n$ for the projection $(x,y)\mapsto(x,-\frac{\partial S}{\partial x}(x,y))$ onto $\mathbb{T}^n\times\mathbb{R}^n$.

E. Garibaldi & P. Thieullen proved in [13] that such strongly minimizing measures exist. Moreover, they are minimizing.

Theorem 3.26. *Let μ be a strongly minimizing measure of a globally positive diffeomorphism of $\mathbb{A}_n$ et let* $\mathrm{supp}\mu$ *be its support. Then*

$$\forall x\in\mathrm{supp}\mu,\ \widetilde{G}_-(x)\le\widetilde{C}_x(\mathrm{supp}\mu)\le\widetilde{G}_+(x).$$

Corollary 3.27. *Let μ be an ergodic strongly minimizing measure of a globally positive diffeomorphism of $\mathbb{A}_n$ all of whose exponents are zero. Then* $\mathrm{supp}\mu$ *is C^1-isotropic almost everywhere.*

References

[1] M.-C. Arnaud, Fibrés de Green et régularité des graphes C^0-Lagrangiens invariants par un flot de Tonelli, *Ann. Henri Poincaré* **9** (2008), no. 5, 881–926.

[2] M.-C. Arnaud, The link between the shape of the Aubry–Mather sets and their Lyapunov exponents, *Ann. Math.* **174-3** (2011), 1571–1601.

[3] M.-C. Arnaud, Green bundles, Lyapunov exponents and regularity along the supports of the minimizing measures, *Ann. Inst. H. Poincaré Anal. Non Linéaire* **29** (2012), no. 6, 989–1007.

[4] M.-C. Arnaud, Lower and upper bounds for the Lyapunov exponents of twisting dynamics: a relationship between the exponents and the angle of Oseledets' splitting, *Ergodic Theory Dynam. Systems* **33** (2013), no. 3, 693–712.

[5] M.-C. Arnaud, Lyapunov exponents for twist maps in high dimension, *Comm. Math. Phys.* preprint arXiv:1409.5203

[6] S. Aubry and P. Y. Le Daeron. The discrete Frenkel-Kontorova model and its extensions. I. Exact results for the ground-states, *Phys. D* **8** (1983), no. 3, 381–422.

[7] P. Bernard. The dynamics of pseudographs in convex Hamiltonian systems, *J.* Amer. Math. Soc. **21** (2008), no. 3, 615–669.

[8] M. L Bialy and R. S. MacKay, Symplectic twist maps without conjugate points, *Israel J. Math.* **141** (2004), 235–247 .
[9] J. Bochi and M. Viana, Lyapunov exponents: how frequently are dynamical systems hyperbolic?, *Modern dynamical systems and applications*, pp. 271–297, Cambridge Univ. Press, Cambridge, 2004.
[10] G. Contreras and R. Iturriaga, Convex Hamiltonians without conjugate points, *Ergodic Theory Dynam. Systems* **19** (1999), no. 4, 901–952.
[11] A. Fathi, *Weak KAM theorems in Lagrangian dynamics*, livre en préparation.
[12] P. Foulon, Estimation de l'entropie des systèmes lagrangiens sans points conjugués *Ann. Inst. H. Poincaré Phys. Théor.* **57** (1992), no. 2, 117–146 .
[13] E. Garibaldi and P. Thieullen, Minimizing orbits in the discrete Aubry-Mather model, *Nonlinearity* **24** (2011), no. 2, 563–611.
[14] C. Golé, Symplectic Twist Maps. Global Variational Techniques. Adv. Ser. Nonlinear Dynam., 18. World Scientific Publishing Co., Inc., River Edge, NJ, 2001. xviii+305 pp.
[15] L. W. Green, A theorem of E. Hopf. *Michigan Math. J* **5** (1958) 31–34.
[16] M. R. Herman, Inégalités "a priori" pour des tores lagrangiens invariants par des difféomorphismes symplectiques. (French) [A priori inequalities for Lagrangian tori invariant under symplectic diffeomorphisms] Inst. Hautes Études Sci. Publ. Math. **70** (1989), 47–101.
[17] M. R. Herman, Dynamics connected with indefinite normal torsion, *Twist Mappings and Their Applications*, 153–182, Springer, New York, 1992.
[18] P. Le Calvez, Les ensembles d'Aubry-Mather d'un difféomorphisme conservatif de l'anneau déviant la verticale sont en général hyperboliques, (French) [The Aubry-Mather sets of a conservative diffeomorphism of the annulus twisting the vertical are hyperbolic in general] *C. R. Acad. Sci. Paris Sér. I Math.* **306** (1988), no. 1, 51–54.
[19] R. S. MacKay, J. D. Meiss and J. Stark, Converse KAM theory for symplectic twist maps, *Nonlinearity* **2** (1989), no. 4, 555–570.
[20] R. Mañé, *Global variational methods in conservative dynamics*, 18 Coloquio Brasileiro de Matematica, IMPA, 1991.
[21] R. Mané, Lagrangian flows: the dynamics of globally minimizing orbits, *Int. Pitman Res. Notes Math. Ser.,* **362**, (1996) 120–131.
[22] J. N. Mather, Existence of quasiperiodic orbits for twist homeomorphisms of the annulus, *Topology* **21** (1982), no. 4, 457–467.
[23] J. N. Mather, Action minimizing invariant measures for positive definite Lagrangian systems, *Math. Z.* **207** (1991), 169–207.
[24] J. Moser, Monotone twist mappings and the calculus of variations, *Ergodic Theory Dynam. Systems* **6-3** (1986), 401–413.

Yonatan Gutman

Takens' embedding theorem with a continuous observable

Abstract: Let (X, T) be a dynamical system where X is a compact metric space and $T : X \to X$ is continuous and invertible. Assume that the Lebesgue covering dimension of X is d. We show that for a generic continuous map $h : X \to [0, 1]$, the $(2d+1)$-delay observation map $x \mapsto \big(h(x), h(Tx), \dots, h(T^{2d}x)\big)$ is an embedding of X inside $[0, 1]^{2d+1}$. This is a generalisation of the discrete version of the celebrated Takens embedding theorem, as proven by Sauer, Yorke and Casdagli to the setting of a continuous observable. In particular there is no assumption on the (lower) box-counting dimension of X which may be infinite.

Keywords: Takens' embedding theorem, continuous observable, time-delay observation map, Lebesgue covering dimension, box-counting dimension.

1 Introduction

Assume a certain physical system, e.g. a certain experimental layout in a laboratory, is modelled by a *dynamical system* (X, T), where $T : X \to X$ represents the state of the system after a certain fixed discrete time interval has elapsed. The possible measurements performed by the experimentalist are modelled by bounded real valued functions $f_i : X \to \mathbb{R}$, $i = 1, \dots, K$ known as *observables*. The actual measurements are performed during a finite time at a discrete rate $t = 0, 1, \dots, N$ starting out in a finite set of initial conditions $\{x_j\}_{j=1}^{L}$. Thus the measurement may be represented by the finite collection of vectors $(f_i(T^k x_j))_{k=0}^{N}$, $i = 1, \dots, K$, $j = 1, \dots, L$. The *reconstruction* problem facing the experimentalist is to characterise (X, T) given this data. Stated in this way the problem is in general not solvable as the obtained data is not sufficient in order to reconstruct (X, T). We, thus, make the unrealistic assumption that the experimentalist has access to $(f_i(T^k x))_{k=0}^{N}$, $i = 1, \dots, K$, $x \in X$. In other words we assume that the experimentalist is able to measure the observable during a finite amount of time, at a discrete rate, starting out with *every* single initial condition. Although this assumption is plainly unrealistic, it enables one, under certain conditions, to solve the reconstruction problem and provide theoretical justification to actual (approximate) procedures used by experimentalists in real life. The first to realise this was F. Takens who proved the famous embedding theorem, now bearing his name.

Theorem. *(Takens' embedding theorem [11, Theorem 1]) Let M be a compact manifold of dimension d. For pairs* (h, T)*, where* $T : M \to M$ *is a* C^2*-diffeomorphism and* $h : M \to \mathbb{R}$ *a* C^2*-function, it is a generic property that the* $(2d+1)$*-delay observation map* $h_0^{2d} : M \to \mathbb{R}^{2d+1}$ *given by*

$$x \mapsto \big(h(x), h(Tx), \ldots, h(T^{2d}x)\big) \tag{1.1}$$

is an embedding, i.e. the set of pairs (h, T) *in* $C^2(M, \mathbb{R}) \times C^2(M, M)$ *for which Equation (1.1) is an embedding is comeagre w.r.t. Whitney* C^2*-topology.*[1]

A key point of the theorem is the possibility to use *one* observable and still be able to achieve embedding through an associated *delay observation map*. Indeed the classical Whitney embedding theorem (see [5, Section 2.15.8]) states that generically a C^2-function $\vec{F} = (F_1, \ldots, F_{2d+1}) : M \to \mathbb{R}^{2d+1}$ is an embedding but this would correspond to the feasibility of measuring $2d+1$ *independent* observables, which is unrealistic for many experimental layouts even if d is small.

A decade after the publication of Takens' embedding theorem it was generalised by Sauer, Yorke and Casdagli in [9]. The generalisation is stronger in several senses. In their theorem the dynamical system is fixed and the embedding is achieved by perturbing solely the observable. This widens the (theoretical) applicability of the theorem but necessitates some assumption about the size of the set of periodic points. Moreover they argue that the concept of (topological) genericity used by Takens is better replaced by a measurable variant of genericity which they call *prevalence*. They also call to attention the fact that in many physical systems the experimentalist tries to characterise a finite dimensional fractal (in particular non-smooth) attractor to which the system converges to, regardless of the initial condition (for sources discussing such systems see [3, 4, 12]). The key point is that although this attractor may be of low fractal dimension, say l , it embeds in phase space in a high-dimensional manifold of dimension, say $n >> l$.[2] As Takens' theorem requires the phase space to be a manifold, it gives the highly inflated number of required measurements $2n + 1$ instead of the more plausible $2l + 1$. Indeed in [9] it is shown that given a C^1-diffeomorphism $T : U \to U$, where $U \subset \mathbb{R}^k$ and a compact $A \subset U$ with *lower box dimension d*, $\underline{\dim}_{\text{box}}(A) = d$, under some technical assumptions on points of low period, it is a prevalent property for $h \in C^1(U, \mathbb{R})$ that the $(2d+1)$-delay observation map $h_0^{2d} : U \to \mathbb{R}^{2d+1}$ is a topological embedding when restricted to A.

In the case of many physical systems, the underlying space in which the finite dimensional attractor arises is infinite dimensional. In [7] Robinson generalised the

1 In [6] Noakes points out that the theorem is also true in the C^1-setting and gives an alternative and more detailed proof. Another detailed and enlightening proof may be found in [10].

2 Notice that, as pointed out in [9, p. 587], it is possible that the minimal dimension of a smooth manifold containing the attractor equals the dimension of the phase space.

previous result to the infinite dimensional context and showed that given a Lipschitz map $T : H \to H$, where H is a Hilbert space and a compact T-invariant set $A \subset H$ with *upper box dimension d*, $\overline{\dim}_{\text{box}}(A) = d$, under some technical assumptions on points of low period, and how well A can be approximated by linear subspaces, it is a prevalent property for Lipschitz maps $h : H \to \mathbb{R}$ that the $(2d+1)$-delay observation map $h_0^{2d} : H \to \mathbb{R}^{2d+1}$ is injective on A.[3] In this work we show that if one is allowed to use continuous (typically non-smooth) observables, then generically one needs even less measurements than previously mentioned in order to reconstruct the original dynamical system. This is achieved by using Lebesgue covering dimension instead of box dimension. We also weaken the invertibility assumption to the more realistic injectivity assumption (see discussion in [12, III.6.2]). We prove the following.

Theorem 1.1. *Let X be a compact metric space and $T : X \to X$ an injective continuous mapping. Assume* $\dim(X) = d$ *and* $\dim(P_n) < 1/2n$ *for all $n \leq 2d$, where* $\dim(\cdot)$ *refers to Lebesgue covering dimension and P_n denotes the set of periodic points of period $\leq n$. Then it is a generic property that the* $(2d+1)$*-delay observation map* $h_0^{2d} : X \to [0,1]^{2d+1}$ *given by*

$$x \mapsto \big(h(x), h(Tx), \ldots, h(T^{2d}x)\big) \tag{1.2}$$

is an embedding, i.e. the set of functions in $C(X,[0,1])$ *for which Equation (1.2) is an embedding is comeagre w.r.t. supremum topology.*

The Lebesgue covering dimension of a compact metric space is always smaller than or equal to the lower box-counting dimension (see [8, Equation 9.1]) and it is not hard to construct compact metric spaces for which the Lebesgue covering dimension is strictly less than the (lower) box-counting dimension, e.g. if C is the Cantor set, then the box dimension of $C^{\mathbb{N}}$ is infinite, whereas the covering dimension is zero. Thus from a theoretical point of view this enables one to reconstruct (using typically a non-smooth observable) dynamical systems with less measurements than were known to suffice previously. Moreover this can be used when the goal of the experiment is to calculate a topological invariant such a *topological entropy*. However I am not certain this result has a bearing on actual experiments. Indeed it has been pointed out to me by physicists that modelling measurements in the lab by smooth functions is realistic, thus non-smooth observables are 'non-accessible' for the experimentalist.

3 Another approach for the infinite dimensional setting is given in [1] with respect to a two-dimensional model of the Navier–Stokes equation. The system has (a typically infinite dimensional) compact *absorbing set*, to which it reaches after a finite and calculable time (depending on the initial condition). It is shown that this set may be embedded in a cubical shift $([0,1])^{\mathbb{Z}}$ through a infinite-delay observation map $x \mapsto (h(x), h(Tx), \ldots)$.

Our result is closely related to a result we published in [2]. In that article it was shown, among other things, that given a finite dimensional topological dynamical system (X, T), where X is a compact metric space with $\dim(X) = d < \infty$ and $T : X \to X$ is a homeomorphism, such that $\dim(P_n) < (1/2)n$ for all $n \le 2d$, (X, T) embeds in the *cubical shift* $([0,1])^Z$, $(X, T) \hookrightarrow (([0,1])^Z, \sigma - \text{shift})$ where the shift action σ is given by $\sigma(x_i)_{i\in Z} = (x_{i+1})_{i\in Z}$. It is not hard to conclude this result from Theorem 1.1 but we are interested in the reverse direction. It would have been possible to rewrite [2] in such a way that Theorem 1.1 follows; however, at the time of its writing we were not aware of the connection to Takens' theorem. Unfortunately a specific part of the proof in [2] uses the fact that $([0,1])^{\mathbb{Z}}$ is infinite dimensional and, therefore, is not straightforwardly adaptable to a proof of Theorem 1.1. In this work we give an alternate and detailed proof of this specific part which is suitable for Theorem 1.1 and indicate how the other parts directly follow from [2]. As mentioned before we only assume $T : X \to X$ is injective and not necessarily a homeomorphism as in [2]. Following Takens we will only deal with the case of *one* observable. The case of several observables follows similarly.

Remark 1.2. Let $(X, (T_t)_{t\in\mathbb{R}})$ be a flow on a compact metric space $X \subset \mathbb{R}^k$ with $\dim(X) = d$, arising from an ordinary differential equation $x = \dot{F}(x)$, where the function $F : X \to \mathbb{R}^k$ obeys the Litschitz condition $\|F(x) - F(y)\| \le L\|x - y\|$. By a theorem of Yorke [13] for any $0 < t < \pi/Ld$ the dynamical system (X, T_t) has no periodic points of order less than $2d + 1$ and, therefore, satisfies the assumptions of Theorem 1.1.

2 Preliminaries

2.1 Dimension

Let $\mathcal{C}$ denote the collection of open (finite) covers of X. Given an open cover $\alpha \in \mathcal{C}$ and a point $x \in X$ we may count the number of elements in α to which x belongs, i.e. $|\{i| x \in U_i\}| = \sum_{U\in\alpha} 1_U(x)$. The *order* of α is essentially defined by maximising this quantity: $\operatorname{ord}(\alpha) = -1 + \max_{x\in X} \sum_{U\in\alpha} 1_U(x)$. Alternatively the *order* of α is the minimal integer n for which any distinct $U_1, U_2, \ldots, U_{n+2} \in \alpha$ obey $\bigcap_{i=1}^{n+2} U_i = \emptyset$. Let $D(\alpha) = \min_{\beta\succ\alpha} \operatorname{ord}(\beta)$ (where β *refines* α, $\beta \succ \alpha$, if for every $V \in \beta$, there is $U \in \alpha$ so that $V \subset U$). The *Lebesgue covering dimension* is defined by $\dim(X) = \sup_{\alpha\in\mathcal{C}} D(\alpha)$.

2.2 Period

For an injective map $T : X \to X$ we define the period of $x \in X$ to be the minimal $p \geq 1$ so that $T^p x = x$. If no such p exists the period is said to be ∞. If the period of x is finite we say x is **periodic**. We denote the set of periodic points in X by P. As T is injective any *preimage of a periodic point is periodic of the same period*. Indeed $T_{|P}$, T restricted to P, is invertible.

2.3 Supremum topology

One defines on $C(X, [0,1])$ the supremum metric $\|\cdot\|_\infty$ given by $\|f - g\|_\infty \triangleq \max_{x\in X} |f(x) - g(x)|$.

3 Proof of the theorem

In this section we prove Theorem 1.1. The proof is closely related to the proof of [2, Theorem 8.1] but unfortunately does not follow directly from it. We, thus, supply the necessary details.

3.1 The Baire category theorem framework

The main tool of the proof is the Baire category theorem. We start with several definitions:

Definition 3.1. A *Baire space* is a topological space where the intersection of countably many dense open sets is dense. By the Baire category theorem $(C(X, [0,1]), \|\cdot\|_\infty)$ is a Baire space. A set in a topological space is said to be *comeagre* or *generic* if it is the complement of a countable union of nowhere dense sets. A set is said to be $\mathbf{G}_\delta$ if it is the countable intersection of open sets. Note that a dense G_δ set is comeagre.

Definition 3.2. Let $K \subset (X \times X) \setminus \Delta$ be a compact set, where $\Delta = \{(x,x) | \, x \in X\}$ is the *diagonal* of $X \times X$ and suppose $h \in C(X, [0,1])$. Denote $h_0^{2d}(x) \triangleq \big(h(x), h(Tx), \ldots, h(T^{2d}x)\big)$. We say that h_0^{2d} is K-compatible if for every $(x,y) \in K$, $h_0^{2d}(x) \neq h_0^{2d}(y)$, or equivalently if for every $(x,y) \in K$, there exists $n \in \{0, 1, \ldots, 2d\}$ so that $h(T^n x) \neq h(T^n y)$. Define:

$$D_K = \{h \in C(X, [0,1]) | \, h_0^{2d} \text{ is } K\text{-compatible}\}$$

In the next section we prove the following key lemma.

Lemma 3.3. *(Main Lemma) One can represent $(X \times X) \setminus \Delta$ as a countable union of compact sets $K_1, K_2, \dots$ such that for all i D_{K_i} is open and dense in $(C(X,[0,1]), \|\cdot\|_\infty)$.*

Proof. [Proof of Theorem 1.1 using Lemma 3.3] As for all i, D_{K_i} is open and dense in $(C(X,[0,1]), \|\cdot\|_\infty)$, we have that $\bigcap_{i=1}^\infty D_{K_i}$ is dense in $(C(X,[0,1]), \|\cdot\|_\infty)$. Any $h \in \bigcap_{i=1}^\infty D_{K_i}$ is K_i-compatible for all i simultaneously and therefore realises an embedding $h_0^{2d} : (X,T) \hookrightarrow [0,1]^{2d+1}$. As a dense G_δ set is comeagre, the above argument shows that the set $\mathcal{A} \subset C(X,[0,1])$ for which $h_0^{2d} : (X,T) \hookrightarrow [0,1]^{2d+1}$ is an embedding is comeagre, or equivalently, that the fact of h_0^{2d} being an embedding is generic in $(C(X,[0,1]), \|\cdot\|_\infty)$. □

It is not hard to see that for every compact $K \subset (X \times X) \setminus \Delta$, D_K is open in $(C(X,[0,1]), \|\cdot\|_\infty)$ (see [2, Lemma A.2]).

3.2 Proof of the main lemma

We write $(X \times X) \setminus \Delta$ as the union of the following three sets: $C_1 = (X \times X) \setminus \big(\Delta \cup (P \times X) \cup (X \times P)\big)$, $C_2 = (P \times P) \setminus \Delta$, $C_3 = \big((X \setminus P) \times P\big) \cup \big(X \times (X \setminus P)\big)$. In words (x,y) (where $x \neq y$) belong to the first, second and third set if both x, y are not periodic, both x, y are periodic and either x or y are periodic but not both, respectively. We then cover each of these sets, $j = 1,2,3$ by a countable union of compact sets $K_1^{(j)}, K_2^{(j)}, \dots$ such that for all i, $D_{K_i^{(j)}}$ is open and dense in $(C(X,[0,1]), \|\cdot\|_\infty)$.

Assume $(x,y) \in C_3$, w.l.o.g. $y \in P$ and $x \notin P$. Denote the period of y by $n < \infty$. Let $t_y = \min\{n-1, 2d\}$. Let H_n be the set of $z \in X$, whose period is n. In other words $H_n = P_n \setminus P_{n-1}$. Notice that H_n is open in P_n and T-invariant. Let U_y be an open set in H_n (but not necessarily open in X) so that $y \in U_y \subset \overline{U}_y \subset H_n$ and $\overline{U}_y \cap T^l\overline{U}_y = \emptyset$ for $l = 1,2,\dots,t_y$. For example, if $d(y, P_{n-1}) = r > 0$, let $0 < \epsilon < r$ small enough so that $U_y = B_\epsilon(y) \cap H_n$ and $\overline{U}_y = \overline{B}_\epsilon(y) \cap P_n = \overline{B}_\epsilon(y) \cap H_n$. As $x \notin P$, the *forward orbit* $\{T^k x\}_{k \geq 0}$ of x is disjoint from P_n. In particular we may choose an open set U_x such that $x \in U_x \subset X \setminus P_n$ (note $X \setminus P_n$ is a T-invariant open set) such that, setting $t_x = 2d$, $\overline{U}_x, T\overline{U}_x, \dots, T^{t_x}\overline{U}_x, \overline{U}_y, T^1\overline{U}_y, \dots, T^{t_y}\overline{U}_y$ are pairwise disjoint. We now define $K_{(x,y)} = \overline{U}_x \times \overline{U}_y$. As X is second-countable, every subspace is a *Lindelöf space*, i.e. every open cover has a countable subcover. For every $n = 1,2,\dots$, H_n can be covered by a countable number of sets of the form U_y. Similarly $X \setminus P$ can be covered by countable number of sets of the form U_x. We can thus choose a countable cover of C_3 by sets of the form $K_{(x,y)}$. We are left with the task of showing that $D_{K_{(x,y)}}$ is dense in $(C(X,[0,1]), \|\cdot\|_\infty)$. Let $\epsilon > 0$. Let $\tilde{f} : X \to [0,1]$ be a continuous function. We will show that there exists a continuous function $f : X \to [0,1]$ so that $\|f - \tilde{f}\|_\infty < \epsilon$ and

f_0^{2d} is $K_{(x,y)}$-compatible. Let α_x and α_y be open covers of $\overline{U}_x$ and $\overline{U}_y$ respectively such that it holds for $j = x, y$ $\max_{W \in \alpha_j, k \in \{0,1,\ldots,t_j\}} \operatorname{diam}(\tilde{f}(T^k W)) < \epsilon/2$ and

$$\operatorname{ord}(\alpha_j) < \frac{t_j + 1}{2} \tag{3.1}$$

For α_x this amounts to $\operatorname{ord}(\alpha_x) \leq d$ which is possible as $\dim(X) = d$ (recall $t_x = 2d$). The same is true for α_y if $t_y \geq 2d$. If $t_y < 2d$, this is possible as by assumption $\dim(\overline{U}_y) \leq \dim(P_{t_y+1}) < (t_j + 1)/2$. For each $W \in \alpha_j$ choose $q_W \in W$ so that $\{q_W\}_{W \in \alpha_j}$ is a collection of distinct points in X. Define $\tilde{v}_W = (\tilde{f}(T^k q_W))_{k=0}^{t_j}$. Notice $t_x \geq t_y$. By [2, Lemma A.9], as Equation (3.1) holds, one can find for $j = x, y$ continuous functions $F_j : \overline{U}_j \to [0,1]^{t_j+1}$, with the following properties:

1. $\forall W \in \alpha_j$, $\|F_j(q_W) - \tilde{v}_W\|_\infty < \epsilon/2$,
2. $\forall z \in \overline{U}_x \cup \overline{U}_y$, $F_j(z) \in \operatorname{co}\{F_j(q_W) | z \in W \in \alpha_j\}$, where $\operatorname{co}(\{v_1, \ldots, v_m\}) \stackrel{\triangle}{=} \{\sum_{i=1}^m \lambda_i v_i | \sum_{i=1}^m \lambda_i = 1, \lambda_i \geq 0\}$,
3. If $x' \in \overline{U}_x$ and $y' \in \overline{U}_y$, then $F_x(x') \neq F_y(y')^{\oplus(2d+1)}$, where $F_y(y')^{\oplus(2d+1)} : \overline{U}_y \to [0,1]^{2d+1}$ is the function given by the formula $[F_y(y')^{\oplus(2d+1)}](k) \stackrel{\triangle}{=} [F_y(y')]\ (k \bmod (t_y + 1))$, $k = 0, 1, \ldots, 2d$.

Let $A = \bigcup_{j=x,y} \bigcup_{k=0}^{t_j} T^k \overline{U}_j$. Define $f' : A \to [0,1]$ ($j = x, y$) by

$$f'_{|T^k \overline{U}_j}(T^k z) = [F_j(z)](k).$$

Fix $z \in \overline{U}_j$ and $k \in \{0, 1, \ldots, t_j\}$. As by Property (2), $f'(T^k z) = [F_j(z)](k) \in \operatorname{co}\{[F_j(q_W)](k) | z \in W \in \alpha_j\}$, we have $|f'(T^k z) - \tilde{f}(T^k z)| \leq \max_{z \in W \in \alpha_j} |[F_j(q_W)](k) - \tilde{f}(T^k z)|$. Fix $W \in \alpha_j$ and $z \in W$. Note $|[F_j(q_W)](k) - \tilde{f}(T^k z)| \leq |[F_j(q_W)](k) - [\tilde{v}_W](k)| + |[\tilde{v}_W](k) - \tilde{f}(T^k z)|$. The first term on the right-hand side is bounded by $\epsilon/2$ by Property (1). As $\operatorname{diam}(\tilde{f}(T^k W)) < \epsilon/2$ and $[\tilde{v}_W](k) = \tilde{f}(T^k q_W)$ we have $|\tilde{f}(T^k q_W) - \tilde{f}(T^k z)| < \epsilon/2$. We finally conclude that $\|f' - \tilde{f}_{|A}\|_\infty < \epsilon$. By an easy application of the Tietze extension theorem (see [2, Lemma A.5]) there is $f : X \to [0,1]$ so that $f_{|A} = f'$ and $\|f - \tilde{f}\|_\infty < \epsilon$. Assume for a contradiction $f_0^{2d}(x') = f_0^{2d}(y')$ for some $(x', y') \in K_{(x,y)}$. This implies that $F_x(x') = (f(x'), \ldots, f(T^{2d} x')) = (f(y'), \ldots, f(T^{2d} y')) = (F_y(y'))^{\oplus(2d+1)}$ which is a contradiction to Property (3).

Unlike the previous case which differs in its treatment from the corresponding case in [2, Theorem 8.1], the cases $(x, y) \in C_1$, $(x, y) \in C_2$ follow quite straightforwardly. Indeed if $(x, y) \in C_1$ (both x and y are not periodic) and in addition the forward orbits of x and y are disjoint then we can use almost verbatim the case $(x, y) \in C_3$. The same is true if $(x, y) \in C_1$ and in addition y belongs to the forward orbit of x, i.e. $y = T^l x$, and $l > 2d$. If $(x, y) \in C_1$, $y = T^l x$ and $l \leq 2d$, then one continues exactly as in Case 2 of [2, Proposition 4.2]. For $(x, y) \in C_2$ one uses [2, Theorem 4.1].

Acknowledgments

I would like to thank Brian Hunt, Ed Ott, Benjamin Weiss and Jim Yorke for helpful discussions. I am grateful to the anonymous reviewer for useful comments. The research was partially supported by the Marie Curie grant PCIG12-GA-2012-334564 and by the National Science Center (Poland) grant 2013/08/A/ST1/00275.

References

[1] Y. Gutman, Embedding topological dynamical systems with periodic points in cubical shifts, To appear in *Ergodic Theory Dynam. Systems* (2016).

[2] Y. Gutman, Mean dimension and Jaworski-type theorems. *Proc. London Math. Soc.* **111** (2005), no. 4, 831–850.

[3] J. K. Hale, Asymptotic behavior of dissipative systems, in: *Mathematical Surveys and Monographs*, Vol. 25, American Mathematical Society, Providence, RI, 1988.

[4] O. Ladyzhenskaya, *Attractors for semigroups and evolution equations*, Lezioni Lincee [Lincei Lectures], Cambridge University Press, Cambridge, 1991.

[5] R. Narasimhan, Analysis on real and complex manifolds, in: *Advanced Studies in Pure Mathematics*, 2nd edition, Vol. 1, Masson & Cie, Éditeur, Paris; North-Holland Publishing Co., Amsterdam-London; American Elsevier Publishing Co., Inc., New York, 1973.

[6] L. Noakes, The Takens embedding theorem. *Internat. J. Bifur. Chaos Appl. Sci. Engrg.* **1** (1991), no. 4, 867–872.

[7] J. C. Robinson, A topological delay embedding theorem for infinite-dimensional dynamical systems, *Nonlinearity* **18** (2005), no. 5, 2135–2143.

[8] J. C. Robinson, Dimensions, embeddings, and attractors, in: *Cambridge Tracts in Mathematics*, Vol. 186, Cambridge University Press, Cambridge, 2011.

[9] T. Sauer, J. A. Yorke and M. Casdagli, Embedology, *J. Statist. Phys.* **65** (1991), nos. 3–4, 579–616.

[10] J. Stark, Delay embeddings for forced systems I.Deterministic forcing, Journal of Nonlinear Science, 9(3): 255–332, 1999.

[11] F. Takens, Detecting strange attractors in turbulence, in: *Dynamical Systems and Turbulence, Warwick 1980 (Coventry, 1979/1980)*, Lecture Notes Math. Vol. 898, Springer, Berlin-New York, 1981, pp. 366–381.

[12] R. Temam, Infinite-dimensional dynamical systems in mechanics and physics, in: *Applied Mathematical Sciences*, 2nd edition, Vol. 68, Springer-Verlag, New York, 1997.

[13] J. A. Yorke, Periods of periodic solutions and the Lipschitz constant, *Proc. Amer. Math. Soc.* **22** (1969), 509–512.

Authors

Ethan Akin
Mathematics Department
The City College
137 Street and Convent Avenue
New York City, NY 10031
USA
ethanakin@earthlink.net

Marie-Claude Arnaud
Université d'Avignon
Laboratoire de Mathématiques d'Avignon
(EA 2151)
84 018 Avignon
France
Marie-Claude.Arnaud@univ-avignon.fr

Joseph Auslander
University of Maryland
Department of Mathematics
College Park, MD 20742-4015
USA
jna@math.umd.edu

Stéphane le Borgne
IRMAR, UMR CNRS 6625
Université de Rennes I
Campus de Beaulieu
35042 Rennes Cedex
France
stephane.leborgne@univ-rennes1.fr

Jean-Pierre Conze
IRMAR, UMR CNRS 6625
Université de Rennes I
Campus de Beaulieu
35042 Rennes Cedex
France
conze@univ-rennes1.fr

Gernot Greschonig
Institute of Discrete Mathematics and Geometry
Vienna University of Technology
Wiedner Hauptstraße 8-10
1040 Vienna
Austria
greschg@fastmail.net

Yonatan Gutman
Institute of Mathematics
Polish Academy of Sciences
ul. Śniadeckich 8
00-656 Warszawa
Poland
y.gutman@impan.pl

Joseph Rosenblatt
Department of Mathematics
University of Illinois at Urbana-Champaign
273 Altgeld Hall, 1409 West Green Street
Urbana, IL 61801
USA
rosnbltt@illinois.edu